

Sequence	Strnd	Orig	ZScore	EScore	Len	Documentation
gb_pat:AX027164	+	27.00	144.97	8.47	32	AX027164 Sequence 24 from Patent
gb_pat:AX082491	+	27.00	142.55	11.54	42	AX082491 Sequence 29 from Patent
gb_pat:AX150169	+	27.00	129.74	59.68	178	AX150169 Sequence 144 from Patent
gb_pl:ATHZ276736	+	27.00	129.35	62.74	186	ATHZ276736 Arabidopsis thaliana
gb_sct:DM8C2T	+	27.00	128.93	66.21	195	D.melanogaster SITS dets
gb_pat:J25411	+	27.00	128.62	68.92	202	J25411 Sequence 7 from patent
gb_sct:HUMSWX3301	+	27.00	128.02	74.38	216	HUMSWX3301 Human chromosome X SITS
gb_sct:AU049240	+	27.00	127.47	79.89	230	AU049240 Rattus norvegicus
gb_ov:RNU07966	+	27.00	126.38	91.85	260	RNU07966 Rattus norvegicus
gb_ov:G283772	+	27.00	125.95	97.09	273	G283772 Gallus microstallite
gb_pat:J22084	+	27.00	125.11	108.09	300	J22084 Sequence 6 from Patent
gb_in:DRSAP528	+	27.00	124.85	111.79	309	DRSAP528 D.pini satellite DNA
gb_in:DMON5AT	-	27.00	124.10	122.97	336	DMON5AT D.montana tandemly repeat
gb_in:DMON1SAT	-	27.00	123.52	132.60	359	DMON1SAT D.montana tandemly repeat
gb_db:AU015522	-	27.00	123.49	132.60	360	AU015522 Uncultured sulfate-reduc
gb_in:AF034222	-	27.00	123.39	134.70	364	AF034222 Gryllus vittatus 12S sm
gb_in:DMON3SAT	-	27.00	123.37	135.12	365	DMON3SAT D.novamexicana tandemly
gb_in:DLUM1SAT	-	27.00	123.32	136.38	367	DLUM1SAT D.lummei tandemly repeat
gb_in:DMON5SAT	-	27.00	123.30	136.38	368	DMON5SAT D.montana tandemly repeat
gb_in:DMON3SAT	-	27.00	123.30	136.38	368	DMON3SAT D.americana tandemly repeat
gb_in:DREX3SAT	-	27.00	123.30	136.38	368	DREX3SAT D.americana tandemly repeat
gb_in:DREX7SAT	-	27.00	123.30	136.38	368	DREX7SAT D.americana tandemly repeat
gb_in:DREX9SAT	-	27.00	123.30	136.38	368	DREX9SAT D.americana tandemly repeat
gb_in:DLIT1SAT	-	27.00	123.27	136.81	369	DLIT1SAT D.littoralis tandemly repeat
gb_in:DLIT2SAT	-	27.00	123.27	136.81	369	DLIT2SAT D.littoralis tandemly repeat
gb_in:DLIT6SAT	-	27.00	123.27	136.81	369	DLIT6SAT D.littoralis tandemly repeat
gb_in:DMON6SAT	-	27.00	123.27	136.81	369	DMON6SAT D.montana tandemly repeat
gb_in:DMON4SAT	-	27.00	123.27	136.81	369	DMON4SAT D.novamexicana tandemly
gb_in:DTEX4SAT	-	27.00	123.27	136.81	369	DTEX4SAT D.americana tandemly repeat
gb_in:DTEX6SAT	-	27.00	123.27	136.81	369	DTEX6SAT D.americana tandemly repeat
gb_in:DVIRK5SAT	-	27.00	123.27	136.81	369	DVIRK5SAT D.virilis tandemly repeat
gb_in:AF034223	-	27.00	123.25	137.23	370	AF034223 Gryllus vittatus 12S sm
gb_in:DAME4SAT	-	27.00	123.25	137.23	370	DAME4SAT D.americana tandemly repeat
gb_in:DAME6SAT	-	27.00	123.25	137.23	370	DAME6SAT D.americana tandemly repeat
gb_in:DLIT3SAT	-	27.00	123.25	137.23	370	DLIT3SAT D.littoralis tandemly repeat
gb_in:DLIT4SAT	-	27.00	123.25	137.23	370	DLIT4SAT D.littoralis tandemly repeat
gb_in:DLIT5SAT	-	27.00	123.25	137.23	370	DLIT5SAT D.littoralis tandemly repeat
gb_in:DLUM2SAT	-	27.00	123.25	137.23	370	DLUM2SAT D.lummei tandemly repeat
gb_in:DLUM3SAT	-	27.00	123.25	137.23	370	DLUM3SAT D.lummei tandemly repeat
gb_in:DLUM4SAT	-	27.00	123.25	137.23	370	DLUM4SAT D.lummei tandemly repeat
gb_in:DLUM5SAT	-	27.00	123.25	137.23	370	DLUM5SAT D.lummei tandemly repeat

Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AX082491

Align seg 1/1 to: AX082491 from: 1 to: 42

1 PheGluPheValGly 5  
|||||  
16 TTCGAGTTTCGTCGC 30

seq\_name: gb\_pat:AX150169

seq\_documentation\_block: 178 bp DNA PAT 08-JUN-2001  
LOCUS AX150169  
DEFINITION Sequence 144 from Patent WO0136685.  
ACCESSION AX150169  
VERSION AX150169.1 GI:14348197  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 178)

AUTHORS Kroes,R.A., Moskal,J.R. and Yamamoto,H.  
TITLE Differential gene expression in cancer  
JOURNAL Patent: WO 0136685-A 144 25-MAY-2001;  
NYXIS Neurotherapies, Inc. (US)

## FEATURES:

source

1..178  
Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 48 a 31 c 26 g 72 t 1 others

## alignment\_scores:

Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AX150169/rev

Align seg 1/1 to reverse of: AX150169 from: 1 to: 178

1 PheGluPheValGly 5  
|||||  
90 TTGGAATTGTGGA 76

seq\_name: gb\_pl:ATH276736

## seq\_documentation\_block:

LOCUS ATH276736 186 bp mRNA PLN 09-MAR-2001  
DEFINITION Arabidopsis thaliana partial mRNA, clone DId1 10A-3a.  
ACCESSION AJ276736  
VERSION AJ276736.1 GI:7263604  
KEYWORDS ORF.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 186)

AUTHORS Vercauteren,I., Van Der Schueren,E., Van Montagu,M. and Gheysen,G.  
TITLE Arabidopsis thaliana genes expressed in the early compatible  
interaction with root-knot nematodes  
Mol. Plant Microbe Interact. 14 (3), 288-299 (2001)

## REFERENCE

2 (bases 1 to 186)

AUTHORS Vercauteren,I.J.R.

TITLE Direct Submission

## JOURNAL

Submitted (11-MAR-2000) Vercauteren I.J.R., Faculteit  
Diergeneeskunde, Vakgroep Parasitologie, University of Ghent,  
Salisburylaan, 133, 9820 Merelbeke, BELGIUM

## FEATURES

source

1..186  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="DId1 10A-3a"  
/tissue\_type="nematode-induced galls"  
<1..>186  
/note="ORF"

BASE COUNT 61 a 28 c 37 g 60 t  
ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x ATH276736

Align seg 1/1 to: ATH276736 from: 1 to: 186

1 PheGluPheValGly 5  
|||||  
81 TTGAGTTTCTAGGA 95

seq\_name: gb\_sts:DM8C2T

## seq\_documentation\_block:

LOCUS DM8C2T 195 bp DNA STS 17-APR-1996  
DEFINITION D. melanogaster STS determined from European Mapping Project  
cosmid, sequence tagged site.

ACCESSION Z32347

VERSION Z32347.1 GI:470259

KEYWORDS STS.

SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 195)

AUTHORS European Drosophila Mapping Consortium.

TITLE Direct Submission

JOURNAL Submitted (08-APR-1994) Michael Ashburner, Department of Genetics,  
Downing St., Cambridge CB2 3EH, England

2 (bases 1 to 195)

AUTHORS European Drosophila Mapping Consortium.

TITLE Direct Submission

JOURNAL Submitted (15-APR-1996) Michael Ashburner, Department of Genetics,  
Downing St., Cambridge CB2 3EH, England

3 (bases 1 to 195)

REFERENCE Updated comments

AUTHORS Madueno,E., Papagiannakis,G., Rimmington,G.A., Saunders,R.D.C.,  
Savakis,C., Siden-Kiamos,I., Skavidis,G., Spanos,L., Treneer,J.,  
Adam,P., Ashburner,M., Benos,P., Bolshakov,V.N., Coulson,D.,  
Glover,D.M., Herrmann,S., Kafatos,F.C., Louis,C., Majerus,T. and  
Moldeli,J.

TITLE A physical map of the X chromosome of Drosophila melanogaster:

cosmid contigs and sequence tagged sites

JOURNAL Genetics 139 (4), 1631-1647 (1995)

MEDLINE 95309678

COMMENT STS\_name = Dm8C2T

clone\_name = 8C2

STS\_from\_promoter = T7

vector\_class = cosmid, Loric 6

origin\_of\_clone = Oregon-R

in\_situ\_site\_primary = 13D

STS\_dbSTS\_AC = 4899

BLAST\_program = BLASTN

database\_searched = EMBL

database\_version = 45.0 and updates till date\_of\_search  
 date\_of\_search = 08-01-1996  
 BLAST\_program = BLASTX  
 database\_searched = SWISSPROT  
 database\_version = 32.0  
 date\_of\_search = 15-12-1995.  
 FEATURES  
 source

BASE COUNT 53 a 40 c 41 g 61 t  
 ORIGIN

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x DM8C2T ..

Align seg 1/1 to: DM8C2T from: 1 to: 195

1 PheGluPheValGly 5  
 |||||  
 71 TTTCGAGTTTGTGGG 85

seq\_name: gb\_pat:125411

seq\_documentation\_block: 202 bp DNA PAT 07-OCT-1996  
 LOCUS 125411  
 DEFINITION Sequence 7 from patent US 5550040.  
 ACCESSION 125411  
 VERSION 125411.1 GI:1605281

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT 52 a 42 c 50 g 58 t  
 ORIGIN

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x 125411 ..

Align seg 1/1 to: 125411 from: 1 to: 202

1 PheGluPheValGly 5  
 |||||  
 149 TTTCGAGTTTGTGGG 163

seq\_name: gb\_sts:HUMSWX3301

seq\_documentation\_block:

LOCUS HUMSWX3301 216 bp DNA STS

DEFINITION Human chromosome X STS SWXD3301, single read, sequence tagged site.

ACCESSION L77880

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

L77880.1 GI:1280295

STS; STS sequence; primer; sequence tagged site.

Homo sapiens DNA.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 216)

Pilla, G., Hughes-Benzle, R.M., Mackenzie, A., Baybayan, P., Chen, E.Y.,

Huber, R., Neri, G., Cao, A., Forabosco, A. and Schlessinger, D.

Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel

overgrowth syndrome

Nature genet. 12 (3), 241-247 (1996)

96172821

Submitted by: David Schlessinger,

Center for Genetics in Medicine,

Washington University School of Medicine, Box 8232 4566 Scott

Avenue

St. Louis, MO 63110, USA

e-mail: david@sequencer.wustl.edu

Primer A: AAGAACTACCAATGCC

Primer B: TTCACCCACAACTCAAAAG

STS size: 75 bp

EXON 3 of GPC3; SWXD3301 was made from L47125.

Location/Qualifiers

1..216

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="Xq26"

32..49

32..106

/standard\_name="GPC3; SWXD3301"

complement(87..106)

BASE COUNT 59 a 49 c 42 g 66 t

ORIGIN

alignment\_scores:

Quality: 27.00 Length: 5

Ratio: 5.400 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-608-892-16 x HUMSWX3301 ..

Align seg 1/1 to: HUMSWX3301 from: 1 to: 216

1 PheGluPheValGly 5

|||||

89 TTTCGAGTTTGTGGG 103

seq\_name: gb\_sts:AU049240

seq\_documentation\_block:

LOCUS AU049240 230 bp DNA STS 20-JAN-2000

DEFINITION Rattus norvegicus, OTSUKA clone, 847e04, microsatellite sequence,

sequence tagged site.

ACCESSION AU049240

VERSION AU049240.1 GI:6722411

KEYWORDS STS.

SOURCE Rattus norvegicus (strain:Brown Norway) liver hepatocyte DNA,

clone:847e04.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (sites)

Watanabe, T.K., Hishigaki, H., Okuno, S., Mizoguchi, A., Oga, K.,

Tsuji, A., Ono, T., Yamasaki, Y., Kanemoto, N., Takahashi, E., Irie, Y.,

Nakamura, Y., Takagi, Y. and Tanigami, A.

The large-scale mapping of rat microsatellite markers

Unpublished (1998)

REFERENCE 2 (bases 1 to 230)

AUTHORS Watanabe, T.K.

```

TITLE      Direct Submission
JOURNAL    Submitted (11-DEC-1998) to the DDBJ/EMBL/GenBank databases. Takeshi
           K Watanabe, Otsuka GEN Research Institute, Otsuka Pharmaceutical
           Co., Ltd; 463-10, Kagasuno, Kawauchi-cho, Tokushima, Tokushima
           771-0192, Japan (E-mail:watanabe@otsuka.gr.jp, Tel:+81-886-65-2888,
           Fax:+81-886-37-1035)
FEATURES   Location/Qualifiers
            source             1..230
                                /organism="Rattus norvegicus"
                                /strain="Brown Norway"
                                /db_xref="taxon:10116"
                                /cell_type="hepatocyte"
                                /clone="847e04"
                                /tissue_type="liver"
                                /note="847e04F=5-CGATCAGCTTCAAGACGCC-3',
                                847e04R=5'-ATCAGGCATGCTTTGT-3'"
BASE COUNT      87 a      67 c      33 g      42 t      1 others
ORIGIN          1..230
                1 PheGluPheValGly 5
                127 TTTGAGTTTGTGGC 113

seq_name: gb_to:RNU07966

seq_documentation_block: 260 bp DNA ROD 24-APR-1994
LOCUS      RNU07966
DEFINITION Rattus norvegicus LEW/N clone DON602A microsatellite sequence.
ACCESSION  U07966
VERSION    U07966.1 GI:474852
KEYWORDS   Norway rat.
SOURCE     Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 260)
           Du,Y., Remmers,E.F., Zha,H. and Wilder,R.L.
           Simple sequence repeats identified in LEW/N rat genomic library
           Unpublished
REFERENCE  2 (bases 1 to 260)
           Du,Y.
           Direct Submission
           Submitted (23-MAR-1994) Ying Du, ARB/NIAMS National Institutes of
           Health 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES   Location/Qualifiers
            source             1..260
                                /organism="Rattus norvegicus"
                                /strain="LEW/N"
                                /db_xref="taxon:10116"
                                /clone="DON602A"
                                /cell_type="hepatocyte"
                                /tissue_type="liver"
                                /clone_lib="AluI/EcoRV/HaeIII/RsaI/SspI restricted genomic
                                DNA"
misc_feature  1..260
              /note="flanking region of microsatellite sequence"
BASE COUNT      85 a      45 c      46 g      84 t
ORIGIN          1..260

alignment_scores:
  Quality: 27.00 Length:
  Ratio: 5.400 Gaps:
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-608-892-16 x AU049240/rev
  Align seg 1/1 to reverse of: AU049240 from: 1 to: 230

  1 PheGluPheValGly 5
  127 TTTGAGTTTGTGGC 113

seq_name: gb_to:RNU07966

seq_documentation_block: 260 bp DNA ROD 24-APR-1994
LOCUS      RNU07966
DEFINITION Rattus norvegicus LEW/N clone DON602A microsatellite sequence.
ACCESSION  U07966
VERSION    U07966.1 GI:474852
KEYWORDS   Norway rat.
SOURCE     Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 260)
           Du,Y., Remmers,E.F., Zha,H. and Wilder,R.L.
           Simple sequence repeats identified in LEW/N rat genomic library
           Unpublished
REFERENCE  2 (bases 1 to 260)
           Du,Y.
           Direct Submission
           Submitted (23-MAR-1994) Ying Du, ARB/NIAMS National Institutes of
           Health 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES   Location/Qualifiers
            source             1..260
                                /organism="Rattus norvegicus"
                                /strain="LEW/N"
                                /db_xref="taxon:10116"
                                /clone="DON602A"
                                /cell_type="hepatocyte"
                                /tissue_type="liver"
                                /clone_lib="AluI/EcoRV/HaeIII/RsaI/SspI restricted genomic
                                DNA"
misc_feature  1..260
              /note="flanking region of microsatellite sequence"
BASE COUNT      85 a      45 c      46 g      84 t
ORIGIN          1..260

alignment_scores:
  Quality: 27.00 Length:
  Ratio: 5.400 Gaps:
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-608-892-16 x RNU07966
  Align seg 1/1 to: RNU07966 from: 1 to: 260

  1 PheGluPheValGly 5
  127 TTTGAGTTTGTGGT 81

seq_name: gb_ov:GG283772

seq_documentation_block: 273 bp DNA VRT 09-JUL-1998
LOCUS      GG283772
DEFINITION G.gallus microsatellite DNA (LEI0199 (-83A02)).
ACCESSION  Z83772
VERSION    Z83772.1 GI:1772528
KEYWORDS   chicken.
SOURCE     Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 273)
           McConnell,S.K.J.
           Direct Submission
           Submitted (02-JAN-1997) McConnell S.K.J., Department of Zoology,
           Adrian Building, University Road, LE1 7RH, Leicester, UK
REFERENCE  2 (bases 1 to 273)
           Dawson,D., McConnell,S., Wardle,A. and Burke,T.A.
           Characterization and mapping of fifteen novel chicken
           microsatellite loci
           Anim. Genet. In Press
FEATURES   Location/Qualifiers
            source             1..273
                                /organism="Gallus gallus"
                                /sub_species="domesticus"
                                /db_xref="taxon:9031"
                                99..134
satellite    /note="microsatellite LEI0199 (-83A02); AC repeat"
BASE COUNT    106 a     59 c     37 g     65 t     6 others
ORIGIN        1..273

alignment_scores:
  Quality: 27.00 Length:
  Ratio: 5.400 Gaps:
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-608-892-16 x GG283772/rev
  Align seg 1/1 to reverse of: GG283772 from: 1 to: 273

  1 PheGluPheValGly 5
  166 TTTGAGTTTGTAGGT 152

seq_name: gb_pat:I22084

seq_documentation_block:
LOCUS      I22084
DEFINITION Sequence 6 from patent US 5525717.
ACCESSION  I22084
VERSION    I22084.1 GI:1602438
KEYWORDS
SOURCE     Unknown.
           Unclassified.
REFERENCE  1 (bases 1 to 300)

```

```

AUTHORS Miyada,C.G. and Born,T.L.
TITLE Support-bound nucleotide probe for neisseria gonorrhoeae
JOURNAL Patent: US 5525717-A 6 11-JUN-1996;
FEATURES
    source
        Location/Qualifiers
            1..300
                /organism="unknown"
BASE COUNT 83 a 58 c 72 g 87 t
ORIGIN

alignment_scores:
    Quality: 27.00 Length: 5
    Ratio: 5.400 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
    US-09-608-892-16 x 122084
    Align seg 1/1 to: 122084 from: 1 to: 300
        1 PheGlupheValGly 5
        |||||||
        155 TTGAGTTTGTGCGA 169

seq_name: gb_in:DPSATPS28

seq_documentation_block:
    LOCUS DPSATPS28 309 bp DNA INV 28-JUL-1999
    DEFINITION D.pini satellite DNA (monomer Ps28).
    ACCESSION X92823
    VERSION X92823.1 GI:5650527
    KEYWORDS satellite.
    SOURCE Diprion pini.
    ORGANISM Diprion pini.
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Hymenoptera; Tenthredinoidea;
        Diprionidae; Diprion.
REFERENCE
    AUTHORS Rouleux-Bonnin,F., Renault,S., Bigot,Y. and Periquet,G.
    TITLE Transcription of four satellite DNA subfamilies in Diprion pini
        (Hymenoptera, Symphyta, Diprionidae)
    JOURNAL Eur. J. Biochem. 238 (3), 752-759 (1996)
    MEDLINE 96300241
REFERENCE
    2 (bases 1 to 309)
    Rouleux-Bonnin,F.
    Direct Submission
    Submitted (03-NOV-1995) F. Rouleux-Bonnin, Inst. de Recherche sur
    la Biologie de l'Insecte, Faculte des Sciences, URA CNRS 1298, Parc
    Grandmont, 37200 Tours, FRANCE
FEATURES
    source
        Location/Qualifiers
            1..309
                /organism="Diprion pini"
                /db_xref="taxon:52634"
    satellite
        1..309
BASE COUNT 91 a 63 c 64 g 91 t
ORIGIN

alignment_scores:
    Quality: 27.00 Length: 5
    Ratio: 5.400 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
    US-09-608-892-16 x DPSATPS28
    Align seg 1/1 to: DPSATPS28 from: 1 to: 309
        1 PheGlupheValGly 5
        |||||||
        55 TTGAGTTTGTGCGG 69

```

```

seq_name: gb_in:DMON2SAT

seq_documentation_block:
    LOCUS DMON2SAT 336 bp DNA INV 22-JAN-1996
    DEFINITION D.montana tandemly repeated satellite DNA (clone pDmon2).
    ACCESSION Z29408
    VERSION Z29408.1 GI:1160419
    KEYWORDS satellite DNA; tandem repeat.
    SOURCE Drosophila montana.
    ORGANISM Drosophila montana.
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
    AUTHORS Bachmann,L.
    TITLE Direct Submission
    JOURNAL Submitted (13-JAN-1994) L. Bachmann, Department of Population
        Genetics, University of Tuebingen, Auf der Morgenstelle 28, 72076
        Tuebingen, Germany
REFERENCE
    2 (bases 1 to 336)
    Heikkinen,E., Launonen,V., Muller,E. and Bachmann,L.
    TITLE The pV8370 BamHI satellite DNA family of the Drosophila virilis
        group and its evolutionary relation to mobile dispersed genetic pdv
        elements
    JOURNAL J. Mol. Evol. 41 (5), 604-614 (1995)
    MEDLINE 96081485
FEATURES
    source
        Location/Qualifiers
            1..336
                /organism="Drosophila montana"
                /strain="1242 (Inari - Finland)"
                /db_xref="taxon:40370"
                /clone="pDmon2"
            50..153
                misc_feature
                /note="significant homology to the pdv mobile dispersed
                    genetic element clone HR2 from position 1425 to 1533;
                    accession number X03936"
            72..195
                misc_feature
                /note="significant homology to the pdv mobile dispersed
                    genetic element clone HR2 from position 658 to 812;
                    accession number X03936"
BASE COUNT 123 a 58 c 49 g 106 t
ORIGIN

alignment_scores:
    Quality: 27.00 Length: 5
    Ratio: 5.400 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
    US-09-608-892-16 x DMON2SAT/rev
    Align seg 1/1 to reverse of: DMON2SAT from: 1 to: 336
        1 PheGlupheValGly 5
        |||||||
        19 TTGAATTCGTAGG 5

seq_name: gb_in:DMON1SAT

seq_documentation_block:
    LOCUS DMON1SAT 359 bp DNA INV 22-JAN-1996
    DEFINITION D.montana tandemly repeated satellite DNA (clone pDmon1).
    ACCESSION Z29407
    VERSION Z29407.1 GI:1160418
    KEYWORDS satellite DNA; tandem repeat.
    SOURCE Drosophila montana.
    ORGANISM Drosophila montana.
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
    1 (bases 1 to 359)
    Bachmann,L.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (13-JAN-1994) L. Bachmann, Department of Population
           Genetics, University of Tuebingen, Auf der Morgenstelle 28, 72076
           Tuebingen, Germany
REFERENCE  2 (bases 1 to 359)
AUTHORS    Heikkinen, E., Launonen, V., Muller, E. and Bachmann, L.
TITLE      The pVb370 BamHI satellite DNA family of the Drosophila virilis
           group and its evolutionary relation to mobile dispersed genetic pdv
           elements
JOURNAL    J. Mol. Evol. 41 (5), 604-614 (1995)
MEDLINE    96081485
FEATURES   Location/Qualifiers
            1..359
            /organism="Drosophila montana"
            /strain="1242 (Inari - Finland)"
            /db_xref="taxon:40370"
            /clone="pDmon1"
            50..149
            /note="significant homology to the pdv mobile dispersed
            genetic element clone HR2 from position 1425 to 1533;
            accession number X03936"
            72..220
            /note="significant homology to the pdv mobile dispersed
            genetic element clone HR2 from position 658 to 812,
            accession number X03936"
BASE COUNT 129 a 60 c 55 g 115 t
ORIGIN
alignment_scores:
    Quality: 27.00      Length: 5
    Ratio: 5.400        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-608-892-16 x DMONISAT/rev ..
Align seg 1/1 to reverse of: DMONISAT from: 1 to: 359
1 PheGluPheValGly 5
|||||
19 TTGAAATCGTAGGG 5

seq_name: gb_ba:AY015522

seq_documentation_block:
LOCUS      AY015522          360 bp      DNA      BCT      29-JUN-2001
DEFINITION Uncultured sulfate-reducing bacterium UMTRadsr617-4 dissimilatory
           sulfite reductase subunit A (dsrA) gene, partial cds.
ACCESSION  AY015522
VERSION    AY015522.1 GI:12667601
KEYWORDS
SOURCE
ORGANISM   uncultured sulfate-reducing bacterium UMTRadsr617-4.
REFERENCE  1 (bases 1 to 360)
           Bacteria; Proteobacteria; delta subdivision; environmental samples.
AUTHORS    Chang, Y.J., Peacock, A.D., Long, P.E., Stephen, J.R., McKinley, J.P.,
           Macnaughton, S.J., Hussain, A.K.M.A., Saxton, A.M. and White, D.C.
TITLE      Diversity and characterization of sulfate-reducing bacteria in
           groundwater at a uranium mill tailings site
JOURNAL    Appl. Environ. Microbiol. 67 (7), 3149-3160 (2001)
MEDLINE    21318708
PUBMED     11425735
REFERENCE  2 (bases 1 to 360)
           Chang, Y.-J., Peacock, A.D., Long, P.E., Stephen, J., McKinley, J.P.,
           Macnaughton, S.J., Hussain, A., Saxton, A.M. and White, D.C.
TITLE      Direct Submission
JOURNAL    Submitted (30-NOV-2000) Center for Environmental Biotechnology,
           University of Tennessee, 10515 Research Drive, Suite 300,
           Knoxville, TN 37932, USA
FEATURES   Location/Qualifiers
            1..360
            /organism="uncultured sulfate-reducing bacterium

```

```

UMTRadsr617-4"
/db_xref="taxon:151106"
/clone="UMTRadsr617-4"
/note="from environmental clones"
<1..>360
/gene="dsrA"
<1..>360
/gene="dsrA"
/transl_table=11
/product="dissimilatory sulfite reductase subunit A"
/protein_id="AAG61205.1"
/db_xref="GI:12667602"
/translation="VICRYSALPEKPPDVAHFHTIRVNQPAQWYTSKALRTLCIDCE
KRGSGLTNMHGSGDIDVFLGTVTNELEPIFSELTENGFDLGGSGDMRTPSCCVGPAP
CEWACYDTIALTDLLTRT"
BASE COUNT 66 a 123 c 100 g 71 t
ORIGIN
alignment_scores:
    Quality: 27.00      Length: 5
    Ratio: 5.400        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-608-892-16 x AY015522/rev ..
Align seg 1/1 to reverse of: AY015522 from: 1 to: 360
1 PheGluPheValGly 5
|||||
213 TTCGAGTTCGTGGT 199

```

---

seq\_name: /SIDS2/qcadata/geneseq/geneseq/NA2000 DAT:AAZ55463



```
seq_documentation_block:
ID_AA255463 standard; DNA: 32 BP.
XX
AC_AA255463;
XX
DT_21-MAR-2000 (first entry)
XX
DE_Neisseria species ORF cloning PCR primer #848.
XX
KW_Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW_antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW_antibacterial; gene therapy; PCR primer; ss.
XX
OS_Synthetic.
OS_Neisseria sp.
XX
PN_WO9957280-A2.
XX
PD_11-NOV-1999.
XX
XX_30-APR-1999; 99WO-US09346.
XX
PR_01-MAY-1998; 98US-0083758.
PR_31-JUL-1998; 98US-0094869.
PR_02-SEP-1998; 98US-0098994.
PR_02-SEP-1998; 98US-0099062.
PR_09-OCT-1998; 98US-0103749.
PR_09-OCT-1998; 98US-0103794.
PR_09-OCT-1998; 98US-0103796.
PR_25-FEB-1999; 98US-0121528.
XX
PA_(CHIR ) CHIRON CORP.
PA_(GENO-) INST GENOMIC RES.
XX
XX_Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX_Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX_Tetteilin H, Venter JC;
XX
XX_WPI; 2000-062150/05.
XX
XX_Novel Neisserial polypeptides predicted to be useful antigens for
XX_vaccines and diagnostics
XX
XX_Example 16; Page 164; 1453pp; English.
XX
XX_AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
XX_represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX_and polypeptides. AA254537 to AA254576 and AA254573 represent
XX_PCR primers used in the exemplification of the present invention. The
XX_polypeptides, the polynucleotides, antibodies and compositions of
XX_the invention can be used as vaccines, as diagnostic reagents, and as
XX_immunogenic compositions. The polypeptides can be used in the
XX_manufacture of medicaments for treating or preventing infection due to
XX_Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX_presence of Neisseria bacteria, or to raise antibodies. They may also
XX_be used to screen for agonists or antagonists, which may themselves
XX_have use as antibacterial agents. The polynucleotides of the invention
XX_may also be used in gene therapy protocols.
XX
XX_Sequence 32 BP; 5 A; 6 C; 9 G; 12 T; 0 other;

alignment_scores:
Quality: 27.00 Length: 5
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x AA255463
Align seg 1/1 to: AA255463 from: 1 to: 32
```

```
1 PheGluPheValGly 5
|||||
11 TTGAAATTGTAGT 25

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF59488

seq_documentation_block:
ID_AA59488 standard; DNA: 42 BP.
XX
AC_AA59488;
XX
DT_03-MAY-2001 (first entry)
XX
DE_L. lactis signal peptide SP310 mutagenesis primer SEQ:29.
XX
KW_Lactococcus lactis; signal peptide; secretion signal; SP310; mutant;
KW_lactic acid bacterium; promoterless promoter reporter; PPR; transposon;
KW_identification; secretion reporter; site directed mutagenesis;
KW_PCR primer; ss.
XX
OS_Lactococcus lactis.
OS_Synthetic.
XX
PN_WO200111060-A2.
XX
PD_15-FEB-2001.
XX
PF_04-AUG-2000; 2000WO-DK00437.
PR_06-AUG-1999; 99DK-0001105.
XX
PA_(BIOT-) BIOTEKNOLOGISK INST.
XX
XX_Ravn P, Madsen SM, Vrang A, Israelsen H, Johnsen MG, Bredmose L;
XX_Arnau J;
XX
XX_WPI; 2001-191547/19.
XX
XX_Constructing a transposon derivative to identify DNA sequence encoding
XX_signal peptide in lactic acid bacteria, involves removing stop codons
XX_in frame with secretion reporter molecule from DNA comprising
XX_transposon
XX
XX_Example 2; Page 26; 62pp; English.
XX
XX_The present invention describes a method for constructing a transposon
XX_derivative for identifying DNA (I) encoding a signal peptide (secretion
XX_signal, SP) in a lactic acid bacterium (e.g. Lactococcus lactis). The
XX_method comprises selecting a transposon (II), including a promoterless
XX_promoter reporter (PPR) gene and a ribosome binding site (RBS), between
XX_its left and right termini (LR) and (RR), deleting a region between LR
XX_and PPR gene to obtain modified DNA that retains its transposability and
XX_its RBS. The present invention also describes: (1) a transposon
XX_derivative (III) for the identification of (I) in a lactic acid
XX_bacterium, comprising (II) without stop codons in the region upstream of
XX_the PPR gene, and a DNA sequence encoding a secretion reporter molecule
XX_comprising at least a part of (III) and (I) that is functional in a
XX_lactic acid bacterium; (3) an isolated DNA sequence (I) coding for SP
XX_derived from SP10, SP13, SP307, SP310 or SP330, or a derivative of any of
XX_the signal peptides having retained signal peptide functionality; (4) a
XX_recombinant plasmid (V) comprising (II) or (I); and (5) a recombinant
XX_bacterium (VI) comprising (I). (III) is useful for identifying and
XX_isolating (I) from a source lactic acid bacteria, by transforming the
XX_bacteria with (III), and selecting from the transformed bacteria, cells
XX_in which the promoterless promoter reporter gene is expressed and the
XX_gene product of the DNA sequence coding for a secretion reporter molecule
XX_is secreted. (VI) is useful for the production of a desired gene product.
XX_AA59460 to AA59499 and AA59428 to AA59472 represent sequences used
XX_in the exemplification of the present invention.
XX
XX_Sequence 42 BP; 8 A; 6 C; 12 G; 16 T; 0 other;
```

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x AAF59488

Align seg 1/1 to: AAF59488 from: 1 to: 42

1 PheGluPheValGly 5  
 16 TTCGAGTTCGTCGCGC 30

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH50815

seq\_documentation\_block:

ID AAH50815 standard; cDNA; 178 BP.

XX

AC AAH50815;

XX 23-AUG-2001 (first entry)

XX Human tumour associated cDNA #144.

KW Human; cancer specific gene expression; gene therapy;  
 KW age related differential expression; ss.

OS Homo sapiens.

XX WO200136685-A2.

XX 25-MAY-2001.

PF 17-NOV-2000; 2000WO-US31809.

PR 17-NOV-1999; 99US-0166056.

XX (NYXI-) NYXIS NEURO THERAPIES INC.

PI Kroes RA, Moskal JR, Yamamoto H;

XX WPI; 2001-355647/37.

XX Novel nucleic acid molecules differentially expressed in brain cancers,  
 PT useful for ascertaining propensity of cell for malignant phenotype or  
 PT ascertaining suitability of anti-neoplastic drug candidate

XX Claim 28; Page 60; 82pp; English.

XX The present invention provides the sequences of 184 cDNA fragments which  
 CC are differentially expressed in cancer cell depending on the age of the  
 CC patient. They can be used to diagnose and identify treatments for  
 CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,  
 CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The  
 CC present sequence is a cancer-associated cDNA of the invention.

XX Sequence 178 BP; 48 A; 31 C; 26 G; 72 T; 1 other;

alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x AAH50815/rev

Align seg 1/1 to reverse of: AAH50815 from: 1 to: 178

1 PheGluPheValGly 5

|||||  
 90 TTGGAATTGTGGA 76

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH81809

seq\_documentation\_block:

ID AAH81809 standard; DNA; 185 BP.

XX

AC AAH81809;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_356 SEQ ID NO:356.

KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.

OS Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

Isolated nucleotide sequences of Neisseria meningitidis which can be  
 used in the diagnosis and treatment of N. meningitidis infection and  
 other Neisserial infections, for example, N.gonorrhoea

Claim 7; Page 1604; 1760pp; English.

The present invention describes methods of obtaining immunogenic  
 proteins from Neisseria genomic sequences. AAH81453 to AAH82414  
 represent specifically claimed Neisseria meningitidis genomic DNA  
 sequences; AAH81260 to AAH81303 and AAH825620 to AAH825663 represent  
 Neisseria DNA sequences and their corresponding proteins; AAH81254 to  
 AAH81259 and AAH81304 to AAH81321 represent PCR primers used in the  
 isolation of Neisseria meningitidis DNA sequences; and AAH81322 to  
 AAH81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 sequences, which are all used in the exemplification of the present  
 invention. The nucleic acid sequences, protein sequences, and antibodies  
 against them, can be used in the manufacture of a composition. The  
 composition can be used as a medicament (or in the manufacture of a  
 medicament) for treating, preventing or diagnosing infection due to  
 Neisserial bacteria. For example, some of the identified proteins could  
 be components of vaccines against Meningococcus B; against all serotypes;  
 and/or against all pathogenic Neisseriae. Identification of sequences  
 from the bacterium will also facilitate production of biological probes,  
 particularly organism-specific probes. Attempts to make efficacious  
 Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 Multivalent vaccines have also been tried but none have successfully  
 overcome antigenic variability. The provision of further, complete  
 sequences may provide an opportunity to identify secreted or surface  
 exposed proteins that may be presumed targets for the immune system and  
 which are not antigenically variable or at least more conserved than  
 other more variable regions.

XX Sequence 185 BP; 30 A; 35 C; 58 G; 62 T; 0 other;

alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AAA81809

Align seg 1/1 to: AAA81809 from: 1 to: 185

1 PheGluPheValGly 5

|||||  
 37 TTCGAGTTTGTGCGA 51

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI21432

seq\_documentation\_block:

ID AAI21432 standard; DNA; 189 BP.

XX AC AAI21432;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #11365 for gene expression analysis in human cervical cell sample.

XX DE Probe; human; microarray; gene expression; cervical epithelial cell.

XX KW cervical cancer; ss.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 25; SEQ ID No 11365; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes

XX CC (SENP). The present sequence is one such probe. The SENPs are derived

XX CC from human HeLa cells. The SENPs can be used to produce a single exon

XX CC microarray, which can be used for measuring human gene expression in a

XX CC sample derived from human cervical epithelial cells. By measuring gene

XX CC expression, the probes are therefore useful in grading and/or staging

XX CC of diseases of the cervix, notably cervical cancer.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 189 BP; 43 A; 48 C; 56 G; 42 T; 0 other;

## alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AAI21432

Align seg 1/1 to: AAI21432 from: 1 to: 189

1 PheGluPheValGly 5

|||||  
 101 TTCGAATTTGTGTC 115

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI46721

seq\_documentation\_block:

ID AAI46721 standard; DNA; 189 BP.

XX AC AAI46721;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #15407 used to measure gene expression in human placenta sample.

XX DE Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-48897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID No 15407; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 189 BP; 43 A; 48 C; 56 G; 42 T; 0 other;

## alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AAI46721

Align seg 1/1 to: AAI46721 from: 1 to: 189

1 PheGluPheValGly 5

|||||  
 101 TTCGAATTTGTGTC 115

seq\_name: /SIDS2/gcgdata/geneseq/NA2001.DAT:AAI07127

seq\_documentation\_block:

ID AAI07127 standard; DNA; 189 BP.

XX

AC AAI07127;

XX

DT 09-OCT-2001 (first entry)

XX

DE Probe #7118 used to measure gene expression in human breast sample.

XX

KW Probe: human; breast disease; breast cancer; development disorder; ss;

XX

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX

OS Homo sapiens.

XX

PN WO200157270-A2.

XX

PD 09-AUG-2001.

XX

PF 29-JAN-2001; 2001WO-US00661.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-476286/51.

XX

XX Novel single exon nucleic acid probe used to measuring gene expression

PT

PT In a human breast.

XX

PS Claim 25; SEQ ID No 7118; 322pp; English.

XX

CC The present invention relates to novel single exon nucleic acid probes.

CC

CC The present sequence is one such probe. The probes are useful for

CC

CC measuring human gene expression in a human breast sample, where the probe

CC

CC hybridises at high stringency to a nucleic acid expressed in the human

CC

CC breast. The probes are useful for predicting, diagnosing, grading,

CC

CC staging, monitoring and prognosing diseases of the human breast.

CC

CC include: breast cancer, disorders of development, inflammatory diseases

CC

CC of the breast, fibrocystic changes, proliferative breast disease and

CC

CC non-carcinoma tumours.

CC

seq\_documentation\_block:

ID AAQ76037 standard; DNA; 202 BP.

XX

AC AAQ76037;

XX

DT 16-JUL-1995 (first entry)

XX

DE N. gonorrhoeae CMT gene.

XX

KW Neisseria gonorrhoeae; cytosine-DNA-methyltransferase; CMT; ss.

XX

OS Neisseria gonorrhoeae.

XX

PN EP630971-A.

XX

PD 28-DEC-1994.

XX

PF 13-JUN-1994; 94EP-0108997.

XX

PR 23-JUN-1993; 93US-0082851.

XX

PR 17-MAR-1994; 94US-0214861.

XX

PA (HOFF ) HOFFMANN LA ROCHE & CO AGF.

XX

PI Purohit AP, Silver SB;

XX

XX WPI; 1995-031607/05.

DR

PT Detection of Neisseria gonorrhoeae and/or Chlamydia trachomatis

PT

XX - simultaneously by a simple, rapid and sensitive technique

XX

PS Disclosure; Fig. 3; 29pp; English.

XX

CC Primers SS01 (given in AAQ76031) and SS02 (AAQ76032) were used for the

CC

CC PCR amplification of a target region (AAQ76037) in the cytosine-DNA-

CC

CC methyltransferase of N. gonorrhoeae. Probe SS06-T5 (AAQ76033) is

CC

CC specific for a region in the amplified sequence, and is used to

CC

CC identify N. gonorrhoeae.

XX

SQ Sequence 202 BP; 52 A; 42 C; 50 G; 58 T; 0 other;

XX

alignment\_scores:

Quality: 27.00 Length: 5

Ratio: 5.400 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

XX

alignment\_block:

US-09-608-892-16 x AAQ76037

XX

Align seg 1/1 to: AAQ76037 from: 1 to: 202

XX

1 PheGluPheValGly 5

|||||

149 TTGAGTTGTCGGA 163

XX

seq\_name: /SIDS2/gcgdata/geneseq/NA2001.DAT:AAH50847

seq\_documentation\_block:

ID AAH50847 standard; cDNA; 202 BP.

XX

AC AAH50847;

XX

DT 23-AUG-2001 (first entry)

XX

DE Human tumour associated cDNA #176.

XX

KW Human; cancer specific gene expression; gene therapy;

XX

KW age related differential expression; ss.

XX

OS Homo sapiens.

XX

seq\_name: /SIDS2/gcgdata/geneseq/NA1995.DAT:AAQ76037

alignment\_scores:

Quality: 27.00 Length: 5

Ratio: 5.400 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

XX

alignment\_block:

US-09-608-892-16 x AAI07127

XX

Align seg 1/1 to: AAI07127 from: 1 to: 189

XX

1 PheGluPheValGly 5

|||||

101 TTCGAATTTGTTGCG 115

```

PN WO200136685-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31809.
XX
XX 17-NOV-1999; 99US-0166056.
XX 17-NOV-1999; 99US-0166106.
XX
XX (NYXI-) NYXIS NEURO THERAPIES INC.
XX
XX Kroes RA, Moskal JR, Yamamoto H;
XX WPI; 2001-355647/37.
XX
XX Novel nucleic acid molecules differentially expressed in brain cancers,
XX useful for ascertaining propensity of cell for malignant phenotype or
XX ascertaining suitability of anti-neoplastic drug candidate -
XX
XX Claim 28; Page 67; 82pp; English.
XX
XX The present invention provides the sequences of 184 cDNA fragments which
XX are differentially expressed in cancer cell depending on the age of the
XX patient. They can be used to diagnose and identify treatments for
XX cancers, particularly brain cancers such as haemangioblastoma, teratoma,
XX haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The
XX present sequence is a cancer-associated cDNA of the invention.
XX
XX Sequence 202 BP; 54 A; 37 C; 32 G; 78 T; 1 other;
SQ

alignment_scores:
    Quality: 27.00      Length: 5
    Ratio: 5.400        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x AAH50847/rev ..

Align seg 1/1 to reverse of: AAH50847 from: 1 to: 202

1 PheGlupheValGly 5
|||||
102 TTGGAATTGTGGA 88

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT.AAC53475

seq_documentation_block:
ID AAC53475 standard; DNA; 244 BP.
XX
XX AAC53475;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 74642.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0123786.
XX 25-MAR-1999; 99US-0126264.
XX

```

```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

```

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150560.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151940.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160988.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x AAC53475 ..  
Align seg 1/1 to: AAC53475 from: 1 to: 244  
1 PheGluPheValGly 5  
|||||  
136 TTGAGTTTGTAGGA 150

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:AAT34990

seq\_documentation\_block:  
ID AAT34990 standard; DNA: 300 BP.  
XX  
AC AAT34990;  
XX  
DT 10-OCT-1996 (first entry)  
XX  
DE N. gonorrhoeae-specific DNA fragment overlapping PCR product.  
XX  
KW Gonorrhoeae; detection; diagnosis; sandwich assay; support; tool;  
KW apparatus; probe; species-specific; ss.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN US5525717-A.  
XX  
PD 11-JUN-1996.  
XX  
PF 09-NOV-1990; 90US-0611528.  
XX  
PR 09-NOV-1990; 90US-0611528.  
PR 25-JUN-1993; 93US-0083946.  
XX  
PA (BEHW ) BEHRINGER AG.  
XX  
PI Born TL, Miyada CG;  
XX

DR WPI; 1996-286454/29.  
 XX Neisseria gonorrhoeae polynucleotide probes - useful for sandwich  
 PT assays and as amplification primer  
 XX  
 PS Example 3; Column 22; 15pp; English.  
 XX  
 CC AAT34988-T34990 are three overlapping PCR products of an 850 nucleotide  
 CC (nt) restriction fragment of Neisseria gonorrhoeae. The three PCR  
 CC products represent bases 1-276, 262-564 and 551-850 of the 850 nt  
 CC fragment (AAT34986). AAT34986 is useful for the production of species-  
 CC specific probes. The probes must be at least 90-100% complementary to  
 CC at least 17 (and pref. 20) contiguous nt of the sequence given. The  
 CC probes are partic. useful as components of a support (e.g. agarose,  
 CC polyacrylamide, nitrocellulose, dextran, chromatographic paper, etc.)  
 CC and esp. when conjugated to a support used in a sandwich hybridisation  
 CC assay, the probes may also be used as primers for N. gonorrhoeae  
 CC nucleic acid. AAT34986 was able to detect 105 out of 106 strains of  
 CC N. gonorrhoeae tested but would not detect any other Neisseria species  
 CC tested e.g. N. meningitidis and N. mucosa. The protein encoded by  
 CC AAT34986 is used to produce species-specific antibodies for the  
 CC detection of N. gonorrhoeae in a sample.  
 XX  
 SQ Sequence 300 BP; 83 A; 58 C; 72 G; 87 T; 0 other;

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Caps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x AAT34990 ..  
 Align seg 1/1 to: AAT34990 from: 1 to: 300  
 1 PheGluPheValGly 5  
 |||||  
 155 TTGTGAGTTGTCGGA 169

seq\_name: /SIDS2/gcgdata/geneseq/NA2000.DAT:AAC29277

seq\_documentation\_block:  
 ID AAC29277 standard; cDNA; 316 BP.  
 XX  
 AC AAC29277;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 33352.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 21-FEB-2000; 2000EP-0200610.  
 XX  
 XX 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX  
 PS Claim 1; SEQ ID 33352; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SQ Sequence 316 BP; 82 A; 50 C; 69 G; 115 T; 0 other;

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Caps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x AAC29277 ..  
 Align seg 1/1 to: AAC29277 from: 1 to: 316  
 1 PheGluPheValGly 5  
 |||||  
 140 TTTGAGTTTGTAGGT 154

seq\_name: /SIDS2/gcgdata/geneseq/NA1998.DAT:AAV25531

seq\_documentation\_block:  
 ID AAV25531 standard; cDNA; 337 BP.  
 XX  
 AC AAV25531;  
 XX  
 DT 28-JUL-1998 (first entry)  
 XX  
 DE UDP-glucose-pyrophosphorylase nucleotide sequence.  
 XX  
 KW Uridine diphosphate glucose pyrophosphorylase gene; transgenic plant;  
 KW resistance; 4-coumarate CoA ligase promoter; 4CL promoter;  
 KW pAX6 binary vector; material re-allocation; glucose; photosynthesis;  
 KW inhibition; photosynthate accumulation; wood production; ss.  
 XX  
 OS Acetobacter xylinum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 24..336  
 FT /\*tag= a  
 FT /product= "UDPG-Ppase"  
 XX  
 PN WO9811240-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 XX 09-SEP-1997; 97WO-CA00631.  
 XX  
 XX 09-SEP-1996; 96US-0707860.  
 XX  
 PA (BCRE-) BC RES INC.  
 XX  
 PI Ellis DD, Gawley JR, Newton CH, Sutton BCS, Xue BG;  
 XX  
 DR WPI; 1998-207399/18.  
 DR P-PSDB; AAW53626.  
 XX

PT Use of Acetobacter xylinum uridine diphosphate glucose  
PT pyrophosphorylase gene - to produce transgenic plants with  
XX increased growth and yield and increased resistance to stress  
PS Example 1; Fig 2; 39pp; English.

XX This is the nucleotide sequence of the uridine diphosphate glucose  
CC pyrophosphorylase gene (UDPG-PPase), which is used in the method of  
CC the invention to create transgenic plants, with increased resistance  
CC to stress. It was used to modify the 4-coumarate CoA ligase (4CL)  
CC promoter. The construct was placed in the pX6 binary vector.  
CC Introduction of the vector (containing the construct) into the plant  
CC causes non-specific re-allocation of materials, e.g. glucose, from  
CC photosynthetic cells to other cells. This removes inhibition on  
CC thus photosynthesis can continue in an uninhibited manner, thereby  
CC increasing growth and yield e.g. wood production in forest trees.  
CC Altering location of cellulose precursor products can also result in  
XX plants with increased resistance to environmental stresses.

SQ Sequence 337 BP; 78 A; 99 C; 89 G; 71 T; 0 other;

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x AAV25531/rev ..

Align seg 1/1 to reverse of: AAV25531 from: 1 to: 337

1 PheGluPheValGly 5  
|||||  
254 TTCAGTTCGTAGGA 240

seq\_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AA03286

seq\_documentation\_block:  
ID AAC03286 standard; cDNA; 350 BP.

AC AAC03286;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3284.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR P-PSDB; AAC03280.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 3284; 71pp + CD-ROM; English.

XX

The present sequence is one of a large number of 5' ESTs derived from  
mRNAs encoding secreted proteins. An ORF has been identified within the  
sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
derived from 30 different tissues. EST sequences usually correspond  
mainly to the 3' untranslated region (UTR) of the mRNA because they are  
often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
well suited for isolating cDNA sequences derived from the 5' ends of  
mRNAs and even in those cases where longer cDNA sequences have been  
obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
mRNAs with intact 5' ends and can therefore be used to obtain full length  
cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
gene therapy and chromosome mapping procedures. They are used to obtain  
upstream regulatory sequences and to design expression and secretion  
vectors.

SQ Sequence 350 BP; 108 A; 69 C; 77 G; 96 T; 0 other;

alignment\_scores:

Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-608-892-16 x AAC03286 ..

Align seg 1/1 to: AAC03286 from: 1 to: 350

1 PheGluPheValGly 5  
|||||  
61 TTTGAGTTTGTTGGA 75





OM of: US-09-608-892-16 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Mar 27, 2002 7:14 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+pn.model -DEV-rlp  
-Q/cgn2\_1/USPTO\_spool/US09608892/runat\_27032002\_11053\_526/app\_query.fasta\_1.63  
-DB-Issued\_Patents\_NA -OFMT-fastap -SUFFIX-rni -GAPOP-12.000  
-GAPOP-4.000 -MINMATCH-0.100 -LOPCL-0.000 -LOPEXT-0.000  
-GAPOP-4.500 -CGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500  
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500  
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-blosom62  
-TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR\_SCORE-pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pfs  
-NORM-ext -MINLEN=0 -MAXLEN=200000000  
-USER-US09608892 @cgn1\_1.65 -NCPU-6 -ICPU-3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS-1

Search information block:

Query: US-09-608-892-16

Query length: 5

Database: Issued\_Patents\_NA.\*

Database sequences: 351203

Database length: 113238999

Search time (sec): 81.430000

score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-214-861-7 +	27.00	120.41	46.71	202	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-083-946-6 +	27.00	117.06	71.78	300	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-915-6 +	27.00	117.06	71.78	300	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-961-083-137 +	27.00	112.70	125.56	502	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-998-418-147 -	27.00	111.84	140.29	556	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-462-965A-1 +	27.00	110.63	163.73	641	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-081-320-4 +	27.00	110.35	169.84	663	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-081-320-15 +	27.00	110.31	170.68	666	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-07-611-528A-1 +	27.00	108.24	222.45	850	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-083-946-1 +	27.00	108.24	222.45	850	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-915-1 +	27.00	108.24	222.45	850	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-906-769-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-906-616-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-639-075A-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-012-431-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-012-692-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-906-613-130 +	27.00	108.19	223.87	855	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-057-929-1 +	27.00	106.32	284.76	1067	
/cgn2_6/ptodata/2/ina/backfiles1.seq:5171850-1 +	27.00	100.11	631.04	2220	Pate
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-081-320-12 +	27.00	98.52	774.24	2680	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-104-073-3 +	27.00	98.20	806.92	2784	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-351-413-10 +	27.00	98.20	806.92	2784	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-025-583-10 +	27.00	98.20	806.92	2784	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-906-395-1 -	27.00	96.89	954.62	3250	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-192-632-1 -	27.00	96.89	954.62	3250	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-710-676-1 -	27.00	96.89	954.62	3250	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-099-902B-1 -	27.00	96.89	954.62	3250	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-099-902B-1 -	27.00	96.89	954.62	3250	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-010-398-14 +	27.00	89.02	2.6e+03	8224	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-366-260-14 +	27.00	89.02	2.6e+03	8224	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-081-320-1 +	27.00	88.51	2.8e+03	8743	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-081-320-26 +	26.00	105.85	302.46	666	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-416-870C-9 +	26.00	98.65	761.25	1558	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-081-320-23 +	26.00	96.50	1.0e+03	2009	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-670-707A-36 +	26.00	86.68	3.5e+03	6402	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-037-601-36 +	26.00	86.68	3.5e+03	6402	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-680-327-2 -	26.00	82.12	6.3e+03	10968	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-228-246-1 -	26.00	82.12	6.3e+03	10968	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-846-111D-15 -	26.00	71.09	2.6e+04	40352	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-697-12 +	24.00	123.18	32.77	30	

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-367-122-45 + 24.00 122.90 33.95 31  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-967-101-142 + 24.00 122.37 36.34 33  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-592-541-142 + 24.00 122.37 36.34 33  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-124-698-142 + 24.00 122.37 36.34 33  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-127-480-142 + 24.00 122.37 36.34 33

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-214-861-7

seq\_documentation\_block:

; Sequence 7, Application US/08214861  
; Patent No. 5550040  
; GENERAL INFORMATION:  
; APPLICANT: Silver, Sheryl B.  
; APPLICANT: Purohit, Ashok P.  
; TITLE OF INVENTION: Method and Reagents for Detection of  
; TITLE OF INVENTION: Neisseria Gonorrhoeae  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patricia S. Rocha  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,861  
; FILING DATE: 17-MAR-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/082,851  
; FILING DATE: 23-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROCHA, PATRICIA S  
; REGISTRATION NUMBER: 31054  
; REFERENCE/DOCKET NUMBER: CD 8698  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-2441  
; TELEFAX: (201) 235-3500  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 202 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; US-08-214-861-7

alignment\_scores:

Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-608-892-16 x US-08-214-861-7 ..

Align seg 1/1 to: US-08-214-861-7 from: 1 to: 202

1 PheGluPheValGly 5

|||||

149 TTGAGTTTGTCGA 163

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-083-946-6

seq\_documentation\_block:

```
; Sequence 6, Application US/08083946
; Patent No. 525717
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
; TITLE OF INVENTION: Gonorrhoeae
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,946
; FILING DATE: 19930625
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/611,528
; FILING DATE: 09-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: Strain #125
; US-08-083-946-6

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-083-946-6
..
Align seg 1/1 to: US-08-083-946-6 from: 1 to: 300

1 PheGluPheValGly 5
|||||
155 TTTGAGTTGTCGGA 169

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-452-915-6

seq_documentation_block:
; Sequence 6, Application US/08452915
; Patent No. 6020461
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
```

```
; TITLE OF INVENTION: Gonorrhoeae
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,515
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/611,528
; FILING DATE: 09-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: Strain #125
; US-08-452-915-6

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-452-915-6
..
Align seg 1/1 to: US-08-452-915-6 from: 1 to: 300

1 PheGluPheValGly 5
|||||
155 TTTGAGTTGTCGGA 169

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-961-083-137

seq_documentation_block:
; Sequence 137, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
```

```

; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; PRIORITY APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1056RP
; US-08-998-416-147

alignment_scores:
    Quality: 27.00      Length: 5
    Ratio: 5.400        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-998-416-147/rev ..

Align seg 1/1 to reverse of: US-08-998-416-147 from: 1 to: 556

1 PheGluPheValGly 5
|||||TTTTTTTTT
208 TTCGAATTCGTTGGG 194

seq_name: /cgn2.6/ptodata/2/ina/5A_COMB.seq:US-08-462-965A-1

seq_documentation_block:
; Sequence 1, Application US/08462965A
; Patent No. 5728546
; GENERAL INFORMATION:
; APPLICANT: Greene Ph.D., John M.
; APPLICANT: Gruber Ph.D., Joachim R.
; APPLICANT: Rosen, Craig R.
; TITLE OF INVENTION: Fibroblast Growth Factor 13
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Oistein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,965A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-439 (PF171)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744

```

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 641 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-462-965A-1

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-608-892-16 x US-08-462-965A-1 ..

Align seg 1/1 to: US-08-462-965A-1 from: 1 to: 641

1 PheCluPheValGly 5

|||||  
573 TTCAGTTGTGGGC 587

seq\_name: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:US-09-081-320-4

seq\_documentation\_block:

; Sequence 4, Application US/09081320  
; Patent No. 6093544

; GENERAL INFORMATION:

; APPLICANT: Gonsalves, Dennis

; APPLICANT: Meng, Baozhong

; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/081,320

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/047,147

; FILING DATE: 20-MAY-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/069,902

; FILING DATE: 17-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/1722

; TELEPHONE: (716) 263-1304

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 663 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-09-081-320-4

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x US-09-081-320-4 ..

Align seg 1/1 to: US-09-081-320-4 from: 1 to: 663

1 PheCluPheValGly 5

|||||  
31 TTGAGTTGTAGGT 45

seq\_name: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:US-09-081-320-15

seq\_documentation\_block:

; Sequence 15, Application US/09081320

; Patent No. 6093544

; GENERAL INFORMATION:

; APPLICANT: Gonsalves, Dennis

; APPLICANT: Meng, Baozhong

; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/081,320

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/047,147

; FILING DATE: 20-MAY-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/069,902

; FILING DATE: 17-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/1722

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1304

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 666 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-09-081-320-15

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-608-892-16 x US-09-081-320-15 ..

```

Align seg 1/1 to: US-09-081-320-15 from: 1 to: 666

1 PheGlupheValgLy 5
|||||
31 TTTGAATTGTGCT 45

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-611-528A-1
seq_documentation_block:
; Sequence 1, Application US/07611528A
; Patent No. 5256536
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/611,528A
; FILING DATE: 19901109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: Strain #125
US-07-611-528A-1

alignment_scores:
Quality: 27.00 Length: 5
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-07-611-528A-1
|||||
1 PheGlupheValgLy 5
705 TTTGAGTTGTGCGA 719

Align seg 1/1 to: US-07-611-528A-1 from: 1 to: 850

1 PheGlupheValgLy 5
|||||
705 TTTGAGTTGTGCGA 719

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-083-946-1
seq_documentation_block:
; Sequence 1, Application US/08452915
; Patent No. 6020461
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria

```

```

; Sequence 1, Application US/08083946
; Patent No. 5525717
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,946
; FILING DATE: 19930625
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/611,528
; FILING DATE: 09-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: Strain #125
US-08-083-946-1

alignment_scores:
Quality: 27.00 Length: 5
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-083-946-1
|||||
1 PheGlupheValgLy 5
705 TTTGAGTTGTGCGA 719

Align seg 1/1 to: US-08-083-946-1 from: 1 to: 850

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-452-915-1
seq_documentation_block:
; Sequence 1, Application US/08452915
; Patent No. 6020461
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria

```

```
; TITLE OF INVENTION: Gonorrhoae
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/611,528
; FILING DATE: 09-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: Strain #125
; US-08-452-915-1

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-452-915-1 ..
Align seg 1/1 to: US-08-452-915-1 from: 1 to: 850

1 PheGluPheValGly 5
|||||
705 TTTCAGTTGTCGCA 719

seq_name: /cgn2_5/ptodata/2/lna/6A_COMB.seq:US-08-906-769-130

seq_documentation_block:
; Sequence 130, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; MOLECULES AND USES THEREOF
```

```
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..759
; OTHER INFORMATION: /note= "At pos. bp 693, change C to
; OTHER INFORMATION: N. At pos. aa 231, substitute Xaa."
; US-08-906-769-130

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-906-769-130 ..
Align seg 1/1 to: US-08-906-769-130 from: 1 to: 855

1 PheGluPheValGly 5
|||||
249 TTTCGATTTGTTGGA 263

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-906-616-130

seq_documentation_block:
; Sequence 130, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 130:
SEQUENCE CHARACTERISTICS:
LENGTH: 855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..759
OTHER INFORMATION: /note= "At pos. bp 693, change C to
OTHER INFORMATION: N. At pos. aa 231, substitute Xaa."
US-08-639-075A-130

alignment_scores:
Quality: 27.00 Length: 5
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x US-08-639-075A-130
Align seg 1/1 to: US-08-639-075A-130 from: 1 to: 855

1 PheGluPheValGly 5
|||||
249 TTCGAATTGTGGA 263

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-012-431-130
seq_documentation_block:
Sequence 130, Application US/09012431
Patent No. 6180383
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 130:
SEQUENCE CHARACTERISTICS:
LENGTH: 855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 1..759
OTHER INFORMATION: /note= "At pos. bp 693, change C to
N. At pos. aa 231, substitute Xaa."
SEQUENCE DESCRIPTION: SEQ ID NO: 130:
US-09-012-431-130

```

```

alignment_scores:
Quality: 27.00 Length: 5
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-608-892-16 x US-09-012-431-130 ..
Align seg 1/1 to: US-09-012-431-130 from: 1 to: 855
1 PheclupheValGly 5
|||||
249 TTCGATTCTGTGGA 263

```

OM of: US-09-608-892-16 to: EST:\* out\_format : pfs

Date: Mar 27, 2002 6:49 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09608892/runat\_27032002\_111052\_494/app\_query.fasta\_1.63  
-DB=EST -QFNT=fastap -SURFIX=st -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09608892@cgn1\_1\_3086 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-608-892-16

Query length: 5

Database: EST\*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1401.500000

score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_est1:AA585229	+	27.00	120.45	1.5e+03	87	AA585229 KTH203 HTCDL1 Homo sapi
gb_est1:AW276168	+	27.00	118.67	1.9e+03	115	AW276168 xq81h09.x1 NCI_CGAP_Br
gb_est1:AW746973	+	27.00	118.29	2.0e+03	122	AW746973 WI_56_E05.g1_A002 Wat
gb_est1:BE950208	-	27.00	118.08	2.1e+03	126	BE950208 uti-W-GE0-ayt-g-02-0-UT
gb_est1:AI114689	-	27.00	118.03	2.1e+03	127	AI146859 oy20g07.s1 Soares_sene
gb_est1:BG27472	-	27.00	117.83	2.1e+03	131	BG267472 1000122D01.x1 1000 - U
gb_est1:BE010091	-	27.00	117.23	2.3e+03	144	BE010091 CM0-BN0180-270400-350-
gb_est1:AI587432	-	27.00	116.55	2.5e+03	160	AI587432 rt50f01.x1 NCI_CGAP_P8
gb_est1:BF955603	-	27.00	116.55	2.5e+03	160	BF955603 RC3-NN0244-181100-011-
gb_est1:KA3052	+	27.00	115.91	2.7e+03	177	KA3052 SW3ICA1048SK Brugia mala
gb_est1:AA192831	+	27.00	115.70	2.8e+03	183	AA192831 zq12c02.s1 Stratagene
gb_est1:BF711485	+	27.00	115.49	2.9e+03	189	BF711485 MI-P-03-abc-c-05-1-UM
gb_gss:AA441555	-	27.00	115.29	2.9e+03	195	AA441555 IM0233F24F Mouse 10kb
gb_est1:AW672301	+	27.00	115.22	3.0e+03	197	AW672301 LGL 358 C07 g1_A002 L4
gb_gss:AQ076753	+	27.00	115.22	3.0e+03	197	AQ076753 CIT-HSP-236165-TF CIT
gb_est1:AA864925	-	27.00	115.06	3.0e+03	202	AA864925 oh44h10.s1 NCI_CGAP_GC
gb_est1:AA010220	-	27.00	114.94	3.1e+03	206	AA010220 z108h03.s1 Soares_feta
gb_est1:D58936	+	27.00	114.82	3.1e+03	210	D58936 HUM516G08B Clontech huma
gb_est1:AI560101	-	27.00	114.52	3.2e+03	220	AI560101 tp12g03.x1 NCI_CGAP_GA
gb_est1:AA942250	-	27.00	114.46	3.3e+03	222	AA942250 LD26353 Sprime LD Dros
gb_est1:BG252413	-	27.00	114.46	3.3e+03	222	BG252413 602366015F1 NIH_MGC-90
gb_est1:AI795487	+	27.00	114.40	3.3e+03	224	AI795487 605010810.Y1 605 - End
gb_est1:AA418273	-	27.00	114.35	3.3e+03	226	AA418273 aj07c07.s1 Soares_para
gb_est1:AA12724	-	27.00	114.35	3.3e+03	226	AA12724 zt67g03.s1 Soares_fetal
gb_est1:AV378647	-	27.00	114.32	3.3e+03	227	AV378647 yf12ell.s1 Soares_fetal
gb_est1:AL366009	+	27.00	114.21	3.4e+03	231	AL366009 MBEA03C02R2 MBEA Medic
gb_est1:BB034303	-	27.00	114.18	3.4e+03	232	BB034303 BB034303 RIKEN full-1e
gb_est1:BB034303	-	27.00	114.10	3.4e+03	235	BB034303 BB034303 RIKEN full-1e
gb_est1:AA911188	-	27.00	114.04	3.5e+03	237	AA911188 ok81c08.s1 NCI_CGAP_K1
gb_est1:AA659365	+	27.00	114.02	3.5e+03	238	AA659365 nt94f01.s1 NCI_CGAP_P1
gb_est1:BE240186	+	27.00	113.96	3.5e+03	240	BE240186 EST404235 MHRP- Medic
gb_est1:W71160	+	27.00	113.96	3.5e+03	240	W71160 me28b04.s1 Soares mouse
gb_gss:CN500WYT	+	27.00	113.94	3.5e+03	241	AL094211 Arabidopsis thaliana g
gb_est1:BB609132	+	27.00	113.88	3.5e+03	243	BB609132 BB609132 RIKEN full-1e
gb_est1:AA919490	-	27.00	113.83	3.5e+03	245	AA919490 RC5-ST0293-201299-011-
gb_est1:BE157557	+	27.00	113.83	3.5e+03	245	BE157557 RC1-HT0375-130400-014-
gb_est1:BF355810	-	27.00	113.80	3.6e+03	246	BF355810 CM1-HT0876-310800-382-
gb_est1:AA879056	-	27.00	113.75	3.6e+03	248	AA879056 nv87h11.s1 NCI_CGAP_P1
gb_est1:BB349640	-	27.00	113.70	3.6e+03	250	BB349640 BB349640 RIKEN full-1e
gb_est1:AA627493	+	27.00	113.70	3.6e+03	250	AA627493 nq46c06.s1 NCI_CGAP_Cd
gb_est1:AA211140	-	27.00	113.68	3.6e+03	251	AA211140 zq87f03.s1 Stratagene

gb\_est2:BI059262 + 27.00 113.68 3.6e+03 251 | BI059262 PM4-GN0511-230401-0  
gb\_est2:Z31117 - 27.00 113.68 3.6e+03 251 | Z31117 MMTES371 Mouse test1  
gb\_gss:AZ101431 - 27.00 113.65 3.6e+03 252 | AZ101431 RPCI-23-465C9-TV RP  
gb\_est1:AV236535 + 27.00 113.60 3.7e+03 254 | AV236535 AV236535 RIKEN full

seq\_name: gb\_est1:AA585229

seq\_documentation\_block:

LOCUS AA585229 87 bp mRNA EST 09-SEP-1997  
DEFINITION KTH203 HTCDL1 Homo sapiens cDNA 5'/3' similar to Mycoplasma DNA  
gyrase, mRNA sequence.  
ACCESSION AA585229  
VERSION AA585229.1 GI:2385117  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1' (bases 1 to 87)  
AUTHORS Sohn,U., Park,D.S., Lee,C.M., Cho,W.K., Ahn,H.J., Lee,M.Y., Hwang  
,M.Y. and Jin,S.W.  
Human HTCDL1 library CDNAS  
Unpublished (1994)

TITLE JOURNAL  
COMMENT Contact: Uik Sohn,D.S.Park,C.M.Lee,W.K.Cho,H.J.Ahn,M.Y.Lee  
,M.Y.Hwang,S.W.Jin  
Laboratory of Molecular Biology  
Kyungpook National University  
Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701,  
Korea

FEATUES source  
1. 87  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HTCDL1"  
/lab\_host="XLI-Blue"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
EcoRI; Poly(A)-mRNA from the 2-year old male fetal  
thymus, oligo(dT) priming, EcoRI cloning in the vector  
pBluescript (Stratagene)."  
BASE COUNT 35 a 15 c 16 g 21 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x AA585229 ..  
Align seg 1/1 to: AA585229 from: 1 to: 87

1 PhEGLuphevaigly 5  
17 TTTGATTTGTTGGT 31

seq\_name: gb\_est1:AW276168

seq\_documentation\_block:

LOCUS AW276168 115 bp mRNA EST 03-JAN-2000  
DEFINITION xq81h09.x1 NCI\_CGAP\_Brn53 Homo sapiens cDNA clone IMAGE:2757089 3',  
mRNA sequence.  
ACCESSION AW276168  
VERSION AW276168.1 GI:6663198  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT

Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 115)  
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)  
Contact: Robert Strausberg, Ph.D.  
Email: cgap@fmail.nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life  
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: polyT not found  
Seq primer: -40UP from Gibco.  
FEATURES  
source  
1..115  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2757089"  
/clone\_lib="NCI-CGAP Brn53"  
/tissue\_type="three pooled meningiomas"  
/lab\_host="DH108"  
/note="Organ: Brain; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies."  
BASE COUNT 34 a 15 c 11 g 55 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x AW746973 rev ..  
Align seg 1/1 to reverse of: AW276168 from: 1 to: 115

1 PheGluPheValGly 5  
|||||  
88 TTTCGATTGTGGA 74

seq\_name: gb\_est1:AW746973

seq\_documentation\_block:  
LOCUS AW746973 122 bp mRNA EST 19-JUL-2000  
DEFINITION WS1.56.E05.g1.A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION AW746973  
VERSION AW746973.1 GI:7660711  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 122)  
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
L.H.  
An EST database from Sorghum: water-stressed plants  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860

Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: T7  
High quality sequence start: 3  
High quality sequence stop: 94  
POLYA=Yes.  
FEATURES  
source  
1..122  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Water-stressed 1 (WS1)"  
/note="Organ: Mix of 5-week old plants on days 7 & 8 after  
water was withheld; Vector: Lambda Zap; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 16 a .18 c 34 g 54 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-608-892-16 x AW746973 ..  
Align seg 1/1 to: AW746973 from: 1 to: 122

1 PheGluPheValGly 5  
|||||  
70 TTTCGATTGTGTCGGGC 84

seq\_name: gb\_est2:BE950208

seq\_documentation\_block:  
LOCUS BE950208 126 bp mRNA EST 04-OCT-2000  
DEFINITION UI-M-CE0-ayt-g-02-0-UI-s1 NIH-BMAP\_Ret3 Mus musculus cDNA clone  
UI-M-CE0-ayt-g-02-0-UI 3', mRNA sequence.  
ACCESSION BE950208  
VERSION BE950208.1 GI:10588874  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 126)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEstemall.nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
retina tissue cDNA Library Preparation. M.B. Soares Lab Clone  
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
GENERICS. It should be noted that Bento Soares is generating a  
small number of additional specialized non-redundant arrays of BMAP  
cDNAs whose availability will be considered under appropriate and

limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

# FEATURES

source  
1..126  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CEO-ayt-g-02-0-UI"  
/clone\_lib="NIH\_BMAP\_Ret3"  
/dev\_stage="6 weeks"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_Ret3 library is derived from mouse retina tissue.  
For a detailed description of the library from which this  
clone was derived, please visit our web site at  
brainest.eng.utoronto.edu.  
TAG\_LIB=NIH\_BMAP\_Ret3  
TAG\_TISSUE=adult-retina  
TAG\_SEQ=GTCCAGCGGCAC"  
BASE COUNT 32 a 28 g 39 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x BE950208/rev ..

Align seg 1/1 to reverse of: BE950208 from: 1 to: 136

1 PheGluPheValGly 5  
|||||

113 TTGGAATTTGTTGGG 99

seq\_name: gb\_est1:A1146859

## seq\_documentation\_block:

LOCUS A1146859 127 bp mRNA EST 29-SEP-1998  
DEFINITION oy20907.s1 Soares\_senescent\_fibroblasts\_NbHSF Homo sapiens cDNA  
clone IMAGE:1666428 3', mRNA sequence.  
ACCESSION A1146859  
VERSION A1146859.1 GI:3674541  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 127)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-re@mail.nih.gov](mailto:cgaps-re@mail.nih.gov)

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](mailto:infoimage.llnl.gov)) for further information.  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 96.  
Location/Qualifiers

# FEATURES

source  
1..127  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1666428"  
/clone\_lib="Soares\_senescent\_fibroblasts\_NbHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker V-type: phagemid; Site\_1: Not I; Site\_2: Eco RI

; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTTCACATCTGAAGTGGAGCGCGCCGATTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo."  
BASE COUNT 38 a 20 c 14 g 55 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 5  
Ratio: 5.400 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x A1146859/rev ..

Align seg 1/1 to reverse of: A1146859 from: 1 to: 127

1 PheGluPheValGly 5  
|||||

83 TTGGAATTTGTTGGA 69

seq\_name: gb\_est2:BG267472

## seq\_documentation\_block:

LOCUS BG267472 131 bp mRNA EST 20-FEB-2001  
DEFINITION 1000122D01.x1 1000 - Unigene I from Maize Genome Project Zea mays  
CDNA, mRNA sequence.  
ACCESSION BG267472  
VERSION BG267472.1 GI:12971418  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
Walbot.V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: [walbot@stanford.edu](mailto:walbot@stanford.edu)  
Plate: 1000122 row: D column: 01.  
Location/Qualifiers

# FEATURES

source  
1..131  
/organism="Zea mays"  
/db\_xref="dbEST:605096F12.x2"  
/db\_xref="taxon:4577"  
/clone\_lib="1000 - Unigene I from Maize Genome Project"  
/note="This library represents the unique ESTs found in  
the first round of EST sequencing at Stanford University  
for the maize genome project. Sequences are present from  
libraries 486, 487, 496, 603, 605, 614, 618, 660, 683  
, 687, 707, and 945. Contigs were assembled using TIGR's  
CAP program and a representative EST from each contig was  
selected for the Unigene set. All singlets were also  
selected."

BASE COUNT 35 a 35 c 26 g 35 t  
ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x BG267472/rev ..  
 Align seg 1/1 to reverse of: BG267472 from: 1 to: 131

1 PheGlupheValGly 5  
 |||||  
 104 TTCGAGTTCGTAGGC 90

seq\_name: gb\_est1:BE010091

seq\_documentation\_block: 144 bp mRNA EST 05-JUN-2000  
 LOCUS BE010091 CM0-BN0180-270400-350-h03 BN0180 Homo sapiens cDNA, mRNA sequence.

DEFINITION BE010091  
 VERSION BE010091.1 GI:8270324

KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 144)  
 Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

## SEQUENCE TAGS

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

## MEDLINE

## COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM0-BN0180-270400-350-h03&t3=2000-04-27&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 4

High quality sequence stop: 92.

## FEATURES

## source

1..144  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0180"  
 /dev\_stage="Adult"

/note="Organ: breast normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 28 a 46 c 29 g 41 t

## ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x BE010091/rev ..

Align seg 1/1 to reverse of: BE010091 from: 1 to: 144

1 PheGlupheValGly 5  
 |||||  
 46 TTCGAGTTTGTGGT 32

seq\_name: gb\_est1:AI587432

seq\_documentation\_block: 160 bp mRNA EST 14-MAY-1999

LOCUS AI587432 tr50f01.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2221753 3',  
 DEFINITION mRNA sequence.

ACCESSION AI587432.1 GI:4573873

VERSION AI587432

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 160)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Life Technologies catalog #: 11548-013

Clone Sequencing by: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 2180 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 105

POLYA-No.

## FEATURES

## source

1..160  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2221753"  
 /clone\_lib="NCI\_CGAP\_Pan1"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SmaI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT 41 a 24 c 19 g 73 t 3 others

## ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-608-892-16 x AI587432/rev ..

Align seg 1/1 to reverse of: AI587432 from: 1 to: 160

1 PheGlupheValGly 5  
 |||||  
 101 TTGGAATTTGTGGA 87

seq\_name: gb\_est2:BF955603

seq\_documentation\_block:

LOCUS BF955603 160 bp mRNA EST 22-JAN-2001

DEFINITION RC5-NN0244-181100-011-F08 NN0244 Homo sapiens cDNA, mRNA sequence.

```

ACCESSION      BF955603
VERSION        BF955603.1  GI:12372878
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 160)
              Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
              Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
              20202663
              Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2704901
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NN0244-
              181100-011-F08&t3=2000-11-18&t4=1)
              Seq primer: puc 18 forward
              High quality sequence start: 9
              High quality sequence stop: 160.
FEATURES      :
source       Location/Qualifiers
              1..160
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="NN0244"
              /dev_stage="Adult"
              /note="Organ: nervous normal; Vector: puc18; Site_1: Sma1;
              Site_2: Sma1; A mini-library was made by cloning products
              derived from ORESTES PCR (O.S. Letters Patent application
              No. 196.716 - Ludwig Institute for Cancer Research)
              profiles into the puc 18 vector. Reverse transcription of
              tissue mRNA and cDNA amplification were performed under
              low stringency conditions."
BASE COUNT   46 a 42 c 34 g 38 t
ORIGIN
alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-608-892-16 x BF955603/rev
  Align seg 1/1 to reverse of: BF955603 from: 1 to: 160
  1 PheGlupheValGly 5
  |||||
  116 TTTGAATTTGCGGT 102
seq_name: gb_est2:N43052
seq_documentation_block:
LOCUS      N43052      177 bp      mRNA      EST      12-DEC-1996
DEFINITION SW31CA1048SK Brugia malayi infective larva cDNA (SAW94WL-BML3)
            Brugia malayi cDNA clone SW31CA1048 5', mRNA sequence.
ACCESSION  N43052
VERSION    N43052.1  GI:1167354
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 183)
            Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.

SOURCE      Brugia malayi.
ORGANISM    Brugia malayi
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia.
            1 (bases 1 to 177)
            Williams,S.A.
            Genes expressed in infective third stage larvae of Brugia malayi
            Unpublished (1995)
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genomesmith.edu
            Seq primer: pbluescript SK.
FEATURES      :
source       Location/Qualifiers
              1..177
              /organism="Brugia malayi"
              /strain="TRS Labs"
              /db_xref="taxon:6279"
              /clone="SW31CA1048"
              /clone_lib="Brugia malayi infective larva cDNA
              (SAW94WL-BML3)"
              /lab_host="XL1-Blue MRF"
              /note="Vector: lambda Unizap XR; Site_1: Ecor I; Site_2:
              Xho I; Lymphatic filarial nematode parasite of humans.
              mRNA was prepared from third stage infective larvae of
              Brugia malayi isolated from mosquitoes 10 days after
              infection and converted to double stranded cDNA using
              reverse transcriptase and oligo(dT) followed by RNase H
              and DNapol I. The library had 1.6 x 10E6 independent
              recombinants and average insert size was 900 base pairs.
              The library was constructed by Weihong Lu. The library is
              available from Dr. S.A. Williams, email genomesmith.edu."
BASE COUNT   49 a 29 c 40 g 59 t
ORIGIN
alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-608-892-16 x N43052
  Align seg 1/1 to: N43052 from: 1 to: 177
  1 PheGlupheValGly 5
  |||||
  107 TTTGAATTTGCGGT 121
seq_name: gb_est1:AA192831
seq_documentation_block:
LOCUS      AA192831      183 bp      mRNA      EST      12-MAR-1998
DEFINITION zql2c02.r1 Stratiagene muscle 937209 Homo sapiens cDNA clone
            IMAGE:629474 5', similar to SW:KPC2_CAEEL P34885 PROTEIN KINASE
            C-LIKE 2 ;, mRNA sequence.
ACCESSION  AA192831
VERSION    AA192831.1  GI:1782237
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 183)
            Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.

```

**TITLE** WashU-NCI human EST Project  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1331 Std Error: 0.00  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 81.

**FEATURES**  
 source  
 1. .183  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5049708"  
 /db\_xref="taxon:9606"  
 /clone="IWAGE:629474"  
 /clone\_lib="Stratagene muscle 937209"  
 /tissue\_type="muscle"  
 /dev\_stage="adult"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skeletal muscle; Vector: pBluescript SK-;  
 Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.  
 Primer: Oligo dT. Skeletal muscle from patient with  
 malignant hyperthermia. Average insert size: 1.0 kb;  
 uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG  
 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'";  
 56 a 50 c 33 g 43 t 1 others

**BASE COUNT**  
**ORIGIN**

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x AA192831/rev ..  
 Align seg 1/1 to reverse of: AA192831 from: 1 to: 183

1 PheGluPheValGly 5  
 |||||  
 51 TTGGAATTGTGGGG 37

seq\_name: gb\_est2:BF711485

seq\_documentation\_block:  
 LOCUS BF711485 189 bp mRNA EST 02-JAN-2001  
 DEFINITION MI-P-03-abc-c-05-1-UM.s1 MI-P-03 Sus scrofa cDNA clone  
 MI-P-03-abc-c-05-1-UM 3', mRNA sequence.  
 ACCESSION BF711485  
 VERSION BF711485.1 GI:12010960  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 189)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 9704477  
 Contact: Tuggle CK  
 Molecular Genetics Laboratory, Department of Animal Science  
 Iowa State University  
 201 Kilgus Hall, Ames, IA 50011-3150, USA  
 Tel: 5152944252  
 Fax: 5152942401  
 Email: cktuggle@iastate.edu

**FEATURES**  
 source  
 1. .189  
 Location/Qualifiers  
 /organism="Sus scrofa"  
 /strain="crossbred"  
 /db\_xref="taxon:9823"  
 /clone="MI-P-03-abc-c-05-1-UM"  
 /clone\_lib="MI-P-03"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-03  
 library is derived from ovary at estrus day 12. For a  
 detailed description of the library from which this clone  
 was derived, please visit our web site at  
 http://pigest.genome.iastate.edu/  
 TAG\_LIB=MI-P-03  
 TAG\_TISSUE=ovary at estrus day 12  
 TAG\_SEQ=TTGTAC"  
 38 a 32 c 55 g 64 t

**BASE COUNT**  
**ORIGIN**

alignment\_scores:  
 Quality: 27.00 Length: 5  
 Ratio: 5.400 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-608-892-16 x BF711485 ..  
 Align seg 1/1 to: BF711485 from: 1 to: 189

1 PheGluPheValGly 5  
 |||||  
 146 TTCGAGTTCGTGGGC 160

seq\_name: gb\_gss:AZ441555

seq\_documentation\_block:  
 LOCUS AZ441555 195 bp DNA GSS 03-OCT-2000  
 DEFINITION 1M0233F24F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0233F24 F, DNA sequence.  
 ACCESSION AZ441555  
 VERSION AZ441555.1 GI:10565568  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 195)  
 Dunn,D., Aoyagi,A., Barber,M., Becsorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 non-normalized ovary at estrus day 12 library cDNA library  
 Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science  
 Research Center, Department of Animal Science, University of  
 Missouri-Columbia, 65211 Clone distribution: clones will be  
 available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=yes.

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0233 row: F column: 24  
 Seq primer: CGTTGTAAACACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 195.  
 Location/Qualifiers

## FEATURES

source

```

1..195
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0233F24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pW42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pW42 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      66 a   40 c   37 g   52 t
ORIGIN

```

```

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-608-892-16 x AZ441555/rev
Align seg 1/1 to reverse of: AZ441555 from: 1 to: 195

```

```

1 PheGlupheValGly 5
|||||
66 TTGGAATTTGTAGGT 52

```

```

seq_name: gb_est1:AW672301

```

```

seq_documentation_block:
LOCUS      AW672301      197 bp      mRNA      EST      19-JUL-2000
DEFINITION LG1_358_C07_g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION  AW672301
VERSION    AW672301.1  GI:7536215
KEYWORDS   EST.
SOURCE     sorghum.
ORGANISM   Sorghum bicolor

```

```

Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 197)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished (2000)
JOURNAL
CONTACT: Cordonnier-Pratt MM

```

Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: T7  
 High quality sequence start: 130  
 High quality sequence stop: 197  
 POLYA=Yes.

## FEATURES

source

```

1..197
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI ; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT      64 a   38 c   58 g   37 t
ORIGIN

```

```

alignment_scores:
  Quality: 27.00      Length: 5
  Ratio: 5.400       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-608-892-16 x AW672301
Align seg 1/1 to: AW672301 from: 1 to: 197

```

```

1 PheGlupheValGly 5
|||||
183 TTGGAATTTGTGGGT 197

```

```

seq_name: gb_gss:AQ076753

```

```

seq_documentation_block:
LOCUS      AQ076753      197 bp      DNA      GSS      20-AUG-1998
DEFINITION CIT-HSP-2361G5.TF CIT-HSP Homo sapiens genomic clone 2361G5, DNA
sequence.
ACCESSION  AQ076753
VERSION    AQ076753.1  GI:3437937
KEYWORDS   GSS.
SOURCE     human.

```

```

ORGANISM   Homo sapiens
Eukaryote: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 197)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Barry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

```

```

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
JOURNAL
COMMENT

```



```

FEATURES                               Location/Qualifiers
  source                                1..197
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="2361G5"
    /clone_lib="CIT-HSP"
    /sex="Male"
    /cell_type="Sperm"
    /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
    HindIII"
BASE COUNT      50 a      37 c      37 g      73 t
ORIGIN

alignment_scores:
  Quality:      27.00      Length:      5
  Ratio:        5.400      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-608-892-16 x AQ076753      ..
Align seg 1/1 to: AQ076753 from: 1 to: 197
  1 PheGluPheValGly 5
  |||||
  81 TTTGAATTTGTCGT 95

```

*Handwritten:*  
 Eric Delaval  
 203.308 = 4698

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2002, 14:53:16 ; Search time 23.65 Seconds

(without alignments)  
15.660 Million cell updates/sec

Title: US-09-608-892-16

Perfect score: 27

Sequence: 1 FEFVG 5.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_1101.\*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
  - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
  - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
  - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
  - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
  - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
  - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
  - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
  - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
  - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
  - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
  - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
  - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
  - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
  - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
  - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
  - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
  - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
  - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
  - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
  - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
  - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	23	AA62221	8-oxo-deoxyguanosine
2	27	100.0	156	AA62221	8-oxo-deoxyguanosine
3	27	100.0	164	AA62221	Fibroblast growth
4	27	100.0	165	AA62221	Human FGF-13 const
5	27	100.0	174	AA62221	Human FGF-13 const
6	27	100.0	175	AA62221	Fibroblast growth
7	27	100.0	185	AA62221	Fibroblast growth
8	27	100.0	185	AA62221	Human FGF-13 const
9	27	100.0	205	AA62221	Amino acid sequenc
10	27	100.0	205	AA62221	FGF-8 homologue PR
11	27	100.0	205	AA62221	Human PRO187 prote

12	27	100.0	205	21	AA62221	Human PRO187 prote
13	27	100.0	205	22	AA62221	Human PRO187 prote
14	27	100.0	205	22	AA62221	Human PRO187 prote
15	27	100.0	205	22	AA62221	Human PRO187 prote
16	27	100.0	205	22	AA62221	Human PRO187 prote
17	27	100.0	211	21	AA62221	Fibroblast growth
18	27	100.0	212	18	AA62221	Human FGF-13 prote
19	27	100.0	212	19	AA62221	Human FGF-13 prote
20	27	100.0	216	19	AA62221	Fibroblast growth
21	27	100.0	216	19	AA62221	Fibroblast growth
22	27	100.0	216	22	AA62221	Human fibroblast g
23	27	100.0	221	20	AA62221	RSPav strain RSP47
24	27	100.0	221	20	AA62221	Peptide #5158 enco
25	27	100.0	233	22	AA62221	Human extracellular
26	27	100.0	239	22	AA62221	Human ORFX ORP2975
27	27	100.0	397	21	AA62221	Human GPC3 protein
28	27	100.0	579	20	AA62221	Human GPC3 protein
29	27	100.0	580	20	AA62221	Novel intestinal o
30	27	100.0	597	14	AA62221	Human PRO5995 poly
31	27	100.0	1024	22	AA62221	Corynebacterium gl
32	27	100.0	95	22	AA62221	Corynebacterium gl
33	27	100.0	95	22	AA62221	Corynebacterium gl
34	27	100.0	95	22	AA62221	C glutamicum prote
35	27	100.0	97	22	AA62221	RSPav strain RSP15
36	27	100.0	221	20	AA62221	Amino acid sequenc
37	27	100.0	310	20	AA62221	H. pylori GHPO 121
38	27	100.0	381	19	AA62221	Raphanus sativus f
39	27	100.0	426	16	AA62221	Consensus sequence
40	27	100.0	426	16	AA62221	Cone snail O-supe
41	27	100.0	426	16	AA62221	Peptide #7392 enco
42	27	100.0	426	16	AA62221	Peptide #10788 enc
43	27	100.0	426	16	AA62221	Arabidopsis thalia
44	27	100.0	426	16	AA62221	Arabidopsis thalia
45	27	100.0	426	16	AA62221	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA62221  
ID AA62221 standard; peptide; 23 AA.

AC AA62221;

DT 26-OCT-1995 (first entry) into word

DE 8-oxo-deoxyguanosine triphosphate decomposition enzyme peptide #4.

KW Human; 8-oxo-deoxyguanosine triphosphate decomposition enzyme;

KX point mutational inhibition; carcinogenesis; cancer; anticancer agent.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 21

FT /note= "Undefined amino acid"

PN JP07023782-A.

PD 27-JAN-1995.

XX 13-JUL-1993; 93JP-0173431.

XX 13-JUL-1993; 93JP-0173431.

XX (SHKJ) SHINGIJUTSU JIGYODAN.

XX WPI; 1995-100944/14.

XX Human 8-oxo-deoxy-guanosine tri-phosphate decomposition enzyme

PT and DNA - useful in the inhibition of point mutation causing

PT cancer

JP  
21 b.b abs

XX PS Claim 1; Page 7; 9pp; Japanese.

CC The sequences given in AAR62218-22 are peptides derived from a human 8-oxo-deoxyguanosine triphosphate decomposition enzyme (see also AAR71800).

CC The enzyme inhibits point mutations caused by a gene which causes mutation and carcinogenesis. The cDNA may be used in the development of diagnosis and the treatment of cancers, and the development of new anticancer agents.

XX SQ Sequence 23 AA;

Query Match 100.0%; Score 27; DB 16; Length 23;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||  
6 fevg 10

Db

RESULT 2

AAR71800

ID AAR71800 standard; Protein; 156 AA.

XX AC AAR71800;

XX DT 26-OCT-1995 (first entry)

XX DE 8-oxo-deoxyguanosine triphosphate decomposition enzyme.

XX KW Human; 8-oxo-deoxyguanosine triphosphate decomposition enzyme; point mutation inhibition; carcinogenesis; cancer; anticancer agent.

XX OS Homo sapiens.

XX PN JP07023782-A.

XX PD 27-JAN-1995.

XX PF 13-JUL-1993; 93JP-0173431.

XX PR 13-JUL-1993; 93JP-0173431.

XX PA (SHKJ) SHINGIJUTSU JIGYODAN.

XX DR WPI: 1995-100944/14.

XX DR N-PSDB; AAQ86263.

XX PT Human 8-oxo-deoxy-guanosine tri-phosphate decomposition enzyme and DNA - useful in the inhibition of point mutation causing cancer

XX PS Claim 3; Page 7-8; 9pp; Japanese.

XX This sequence represents a human 8-oxo-deoxyguanosine triphosphate decomposition enzyme. The enzyme inhibits point mutations caused by a gene which causes mutation and carcinogenesis. The cDNA may be used in the development of diagnosis and the treatment of cancers, and the development of new anticancer agents.

XX SQ Sequence 156 AA;

Query Match 100.0%; Score 27; DB 16; Length 156;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||  
72 fevg 76

Db

RESULT 3

AAW70331

ID AAW70331 standard; Protein; 164 AA.

XX AC AAW70331;

XX DT 18-NOV-1998 (first entry)

XX DE Fibroblast growth factor-13 deletion variant 1.

XX KW Fibroblast growth factor-13; FGF-13; cell proliferation; angiogenesis; revascularisation; ischaemia; limb regeneration; wound; neuronal growth; stroke; chondrocyte; skin ageing; haematopoietic cell; bone marrow cell; cytokine; neoplastic cell; tumour; cataract; epithelial lens cell.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key

XX FT Region

XX FT Location/Qualifiers

1..164

/note= "represents residues 43-207 of the wild-type FGF-13 protein (AAW70330)"

XX PN WO9823749-A1.

XX PD 04-JUN-1998.

XX PF 21-NOV-1997; 97WO-US20548.

XX PR 04-DEC-1996; 96US-0031575.

XX PR 27-NOV-1996; 96US-0031969.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Greene JM, Gruber JR, Rosen CA;

XX DR WPI: 1998-322729/28.

XX DR N-PSDB; AAV33224.

XX PT New isolated fibroblast growth factor-13 - used to develop products for treating e.g. cardiovascular conditions, wounds, neurodegenerative conditions, tumours or hyper-vascular diseases

XX PS Example 7; Page 80; 132pp; English.

XX The present sequence represents the fibroblast growth factor-13 (FGF-13) deletion variant 1 (5' delta 42/3' delta 9) protein. This mutant FGF-13 protein has the first 42 N-terminal residues and the last 9 residues at the C-terminal deleted when compared with the wild-type FGF-13 protein (AAW70330). This deletion variant protein is given as an example of a mutant protein which is capable of retaining its biological activity. The invention claims for a FGF-13 encoding DNA sequence (AAV33211) isolated from a cDNA library derived from human ovarian cancer tissue. The FGF-13 protein belongs to the fibroblast growth factor family and it is known to stimulate cellular proliferation. FGF-13 protein is claimed to be useful for treating or stimulating revascularisation of ischemic tissues, for stimulating angiogenesis and limb regeneration, for treating wounds due to injuries, burns and ulcers. The FGF-13 protein is also claimed to be useful for stimulating neuronal growth and for treating and preventing neuronal damage associated with stroke, for stimulating chondrocyte growth, for preventing skin ageing due to sunburn and for stimulating growth and differentiation of haematopoietic cells and bone marrow cells when used in combination with other cytokines. FGF-13 antagonists are claimed to be useful for inhibiting the cell growth and proliferation effects of the products on neoplastic cells and tissues, i.e. stimulation of angiogenesis of tumours, for preventing the proliferation of epithelial lens cells after extracapsular cataract surgery or for preventing the growth of scar tissue during wound healing.

SQ Sequence 164 AA;  
 Query Match 100.0%; Score 27; DB 19; Length 164;  
 Best Local Similarity 100.0%; Pred. NO. 48;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FEFVG 5  
 Db 152 fefvg 156  
 |||||  
 + RESULT 4  
 AAB50282  
 ID AAB50282 standard; Protein; 165 AA.  
 XX AC AAB50282;  
 XX DT 20-MAR-2001 (first entry)  
 XX DE Human FGF-13 construct protein SEQ ID NO: 28.  
 XX KW Human: fibroblast growth factor 13; FGF-13; gene therapy; cancer;  
 KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
 KW wound healing; neurological disease; infection; tissue regeneration.  
 XX OS Homo sapiens.  
 XX PN WO200071567-A2.  
 XX PD 30-NOV-2000.  
 XX PF 04-MAY-2000; 2000WO-US40080.  
 XX PR 06-MAY-1999; 99US-0132923.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Melder R, Duan RD, Alderson R, Rosen CA, Greene JM, Gruber JR;  
 XX DR WPI; 2001-025133/03.  
 XX Nucleic acids encoding human fibroblast growth factor 13 and their  
 PT corresponding proteins, useful for gene therapy -  
 XX PS Example 9; Page 277-278; 282pp; English.  
 XX CC The present invention provides the protein and coding sequences for human  
 CC fibroblast growth factor 13 (FGF-13). These can be used in the prevention  
 CC and treatment of hyperproliferative disorders, cardiovascular disorders,  
 CC cancers, neurological diseases, infections, in the promotion of wound  
 CC healing and tissue regeneration and in the prevention of angiogenesis.  
 XX SQ Sequence 165 AA;  
 Query Match 100.0%; Score 27; DB 22; Length 165;  
 Best Local Similarity 100.0%; Pred. NO. 48;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FEFVG 5  
 Db 153 fefvg 157  
 |||||  
 + RESULT 5  
 AAB50283  
 ID AAB50283 standard; Protein; 174 AA.  
 XX AC AAB50283;  
 XX DT 20-MAR-2001 (first entry)  
 XX DE Human FGF-13 construct protein SEQ ID NO: 32.  
 XX KW Human: fibroblast growth factor 13; FGF-13; gene therapy; cancer;  
 KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
 KW wound healing; neurological disease; infection; tissue regeneration.  
 XX OS Homo sapiens.  
 XX PN WO200071567-A2.  
 XX PD 30-NOV-2000.  
 XX PF 04-MAY-2000; 2000WO-US40080.  
 XX PR 06-MAY-1999; 99US-0132923.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Melder R, Duan RD, Alderson R, Rosen CA, Greene JM, Gruber JR;  
 XX DR WPI; 2001-025133/03.  
 XX Nucleic acids encoding human fibroblast growth factor 13 and their  
 PT corresponding proteins, useful for gene therapy -  
 XX PS Example 9; Page 278-279; 282pp; English.  
 XX CC The present invention provides the protein and coding sequences for human  
 CC fibroblast growth factor 13 (FGF-13). These can be used in the prevention  
 CC and treatment of hyperproliferative disorders, cardiovascular disorders,  
 CC cancers, neurological diseases, infections, in the promotion of wound  
 CC healing and tissue regeneration and in the prevention of angiogenesis.  
 XX SQ Sequence 174 AA;  
 Query Match 100.0%; Score 27; DB 22; Length 174;  
 Best Local Similarity 100.0%; Pred. NO. 50;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FEFVG 5  
 Db 153 fefvg 157  
 |||||  
 + RESULT 6  
 AAW70332  
 ID AAW70332 standard; Protein; 175 AA.  
 XX AC AAW70332;  
 XX DT 18-NOV-1998 (first entry)  
 XX DE Fibroblast growth factor-13 deletion variant 2.  
 XX KW Fibroblast growth factor-13; FGF-13; cell proliferation; angiogenesis;  
 KW revascularisation; ischaemia; limb regeneration; wound;  
 KW neuronal growth; stroke; chondrocyte; skin ageing; haematopoietic cell;  
 KW bone marrow cell; cytokine; neoplastic cell; tumour; cataract;  
 KW epithelial lens cell.  
 XX OS Synthetic.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT Region 1..175  
 FT /note= "represents residues 43-216 of the wild-type  
 FT FGF-13 protein (AAW70330)"  
 XX PN WO9823749-A1.  
 XX PD 04-JUN-1998.  
 XX

PF 21-NOV-1997; 97WO-US20548.  
 XX  
 PR 04-DEC-1996; 96US-0031575.  
 XX 27-NOV-1996; 96US-0031969.  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Greene JM, Gruber JR, Rosen CA;  
 XX  
 DR WPI; 1998-322729/28.  
 XX N-PSDB; AAV33225.  
 DR  
 XX New isolated fibroblast growth factor-13 - used to develop products  
 PT for treating e.g. cardiovascular conditions, wounds,  
 PT neurodegenerative conditions, tumours or hyper-vascular diseases  
 XX  
 PS Example 7; Page 80; 132pp; English.  
 XX  
 CC The present sequence represents the fibroblast growth factor-13  
 CC (FGF-13) deletion variant 2 (5' delta 42/3' full) protein. This  
 CC mutant FGF-13 protein has the first 42 N-terminal residues deleted  
 CC whilst retaining the C-terminal residues when compared with the  
 CC wild-type FGF-13 protein (AAW70330). This deletion variant protein  
 CC is given as an example of a mutant protein which is capable of  
 CC retaining its biological activity. The invention claims for a FGF-13  
 CC encoding DNA sequence (AAV33211) isolated from a cDNA library derived  
 CC from human ovarian cancer tissue. The FGF-13 protein belongs to the  
 CC fibroblast growth factor family and it is known to stimulate cellular  
 CC proliferation. FGF-13 protein is claimed to be useful for treating or  
 CC stimulating revascularisation of ischemic tissues, for stimulating  
 CC angiogenesis and limb regeneration, for treating wounds due to  
 CC injuries, burns and ulcers. The FGF-13 protein is also claimed to be  
 CC useful for stimulating neuronal growth and for treating and preventing  
 CC neuronal damage associated with stroke, for stimulating chondrocyte  
 CC growth, for preventing skin ageing due to sunburn and for stimulating  
 CC growth and differentiation of haematopoietic cells and bone marrow  
 CC cells when used in combination with other cytokines. FGF-13 antagonists  
 CC are claimed to be useful for inhibiting the cell growth and  
 CC proliferation effects of the products on neoplastic cells and tissues,  
 CC i.e. stimulation of angiogenesis of tumours, for preventing the  
 CC proliferation of epithelial lens cells after extracapsular cataract  
 CC surgery or for preventing the growth of scar tissue during wound  
 CC healing.  
 XX  
 SQ Sequence 175 AA;  
 Query Match 100.0%; Score 27; DB 19; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PEFVG 5  
 Db |||||  
 154 fefvg 158  
 RESULT 7  
 AAW70333  
 ID AAW70333 standard; Protein; 185 AA.  
 XX  
 AC AAW70333;  
 XX  
 XX 18-NOV-1998 (first entry)  
 DT  
 XX Fibroblast growth factor-13 deletion variant 3.  
 DE  
 XX Fibroblast growth factor-13; FGF-13; cell proliferation; angiogenesis;  
 KW revascularisation; ischaemia; limb regeneration; wound;  
 KW neuronal growth; stroke; chondrocyte; skin ageing; haematopoietic cell;  
 KW bone marrow cell; cytokine; neoplastic cell; tumour; cataract;  
 KW epithelial lens cell.  
 XX  
 OS Synthetic.

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "This residue was introduced into the DNA  
 FT sequence encoding the FGF-13 variant 3 protein"  
 FT  
 FT Region 2..185  
 FT /note= "represents residues 24-207 of the wild-type  
 FT FGF-13 protein (AAW70330)"  
 XX  
 PN WO9823749-A1.  
 XX  
 PD 04-JUN-1998.  
 XX  
 PF 21-NOV-1997; 97WO-US20548.  
 XX  
 PR 04-DEC-1996; 96US-0031575.  
 XX 27-NOV-1996; 96US-0031969.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PA Greene JM, Gruber JR, Rosen CA;  
 XX  
 PI  
 XX  
 DR WPI; 1998-322729/28.  
 DR N-PSDB; AAV33226.  
 XX  
 PT New isolated fibroblast growth factor-13 - used to develop products  
 PT for treating e.g. cardiovascular conditions, wounds,  
 PT neurodegenerative conditions, tumours or hyper-vascular diseases  
 XX  
 PS Example 7; Page 81; 132pp; English.  
 XX  
 CC The present sequence represents the fibroblast growth factor-13  
 CC (FGF-13) deletion variant 3 (5' delta 23/3' full) protein. This  
 CC mutant FGF-13 protein has the first 23 N-terminal residues deleted  
 CC whilst retaining the C-terminal residues when compared with the  
 CC wild-type FGF-13 protein (AAW70330). This deletion variant protein  
 CC is given as an example of a mutant protein which is capable of  
 CC retaining its biological activity. The invention claims for a FGF-13  
 CC encoding DNA sequence (AAV33211) isolated from a cDNA library derived  
 CC from human ovarian cancer tissue. The FGF-13 protein belongs to the  
 CC fibroblast growth factor family and it is known to stimulate cellular  
 CC proliferation. FGF-13 protein is claimed to be useful for treating or  
 CC stimulating revascularisation of ischemic tissues, for stimulating  
 CC angiogenesis and limb regeneration, for treating wounds due to  
 CC injuries, burns and ulcers. The FGF-13 protein is also claimed to be  
 CC useful for stimulating neuronal growth and for treating and preventing  
 CC neuronal damage associated with stroke, for stimulating chondrocyte  
 CC growth, for preventing skin ageing due to sunburn and for stimulating  
 CC growth and differentiation of haematopoietic cells and bone marrow  
 CC cells when used in combination with other cytokines. FGF-13 antagonists  
 CC are claimed to be useful for inhibiting the cell growth and  
 CC proliferation effects of the products on neoplastic cells and tissues,  
 CC i.e. stimulation of angiogenesis of tumours, for preventing the  
 CC proliferation of epithelial lens cells after extracapsular cataract  
 CC surgery or for preventing the growth of scar tissue during wound  
 CC healing.  
 XX  
 SQ Sequence 185 AA;  
 Query Match 100.0%; Score 27; DB 19; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PEFVG 5  
 Db |||||  
 173 fefvg 177  
 RESULT 8  
 AAB50284  
 ID AAB50284 standard; Protein; 185 AA.

XX AAB50284;  
AC DT 20-MAR-2001 (first entry)  
XX DE XX  
XX DE Human FGF-13 construct protein SEQ ID NO: 36.  
XX DE  
KW Human; fibroblast growth factor 13; FGF-13; gene therapy; cancer;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW wound healing; neurological disease; infection; tissue regeneration.  
XX OS Homo sapiens.  
XX XX  
XX PN WO200071567-A2.  
XX XX  
XX PD 30-NOV-2000.  
XX XX  
XX PF 04-MAY-2000; 2000WO-US40080.  
XX XX  
XX PR 06-MAY-1999; 98US-0132923.  
XX XX  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX XX  
XX PI Melder R, Duan RD, Alderson R, Rosen CA, Greene JM, Gruber JR;  
XX XX  
XX DR WPI; 2001-025133/03.  
XX XX  
XX PT Nucleic acids encoding human fibroblast growth factor 13 and their  
XX PT corresponding proteins, useful for gene therapy -  
XX XX  
XX PS Example 9; Page 280; 282pp; English.  
XX XX  
XX CC The present invention provides the protein and coding sequences for human  
XX CC fibroblast growth factor 13 (FGF-13). These can be used in the prevention  
XX CC and treatment of hyperproliferative disorders, cardiovascular disorders,  
XX CC cancers, neurological diseases, infections, in the promotion of wound  
XX CC healing and tissue regeneration and in the prevention of angiogenesis.  
XX SQ Sequence 185 AA;

Query Match 100.0%; Score 27; DB 22; Length 185;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5  
| | | | |  
Db 173 fefvg 177

RESULT 9  
AAY13348  
ID AAY13348 standard; Protein; 205 AA.  
XX AC AAY13348;  
XX AC  
XX DT 25-JUN-1999 (first entry)  
XX DE Amino acid sequence of protein PRO187.  
XX XX  
XX KW Secreted protein; transmembrane protein; human; enterocolitis;  
XX KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
XX KW congenital microvillus atrophy; skin disease; cell growth;  
XX KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
XX KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
XX KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
XX KW anti-thrombotic; wound healing; tissue repair.  
XX OS Homo sapiens.  
XX XX  
XX PN WO9914328-A2.  
XX XX  
XX PD 25-MAR-1999.

XX PF 16-SEP-1998; 98WO-US19330.  
XX XX  
XX PR 25-NOV-1997; 97US-0066840.  
XX PR 17-SEP-1997; 97US-0059113.  
XX PR 17-SEP-1997; 97US-0059115.  
XX PR 17-SEP-1997; 97US-0059117.  
XX PR 17-SEP-1997; 97US-0059119.  
XX PR 17-SEP-1997; 97US-0059121.  
XX PR 17-SEP-1997; 97US-0059122.  
XX PR 17-SEP-1997; 97US-0059124.  
XX PR 18-SEP-1997; 97US-0059263.  
XX PR 18-SEP-1997; 97US-0059266.  
XX PR 15-OCT-1997; 97US-0062125.  
XX PR 17-OCT-1997; 97US-0062285.  
XX PR 17-OCT-1997; 97US-0062287.  
XX PR 21-OCT-1997; 97US-0063486.  
XX PR 24-OCT-1997; 97US-0062814.  
XX PR 24-OCT-1997; 97US-0062816.  
XX PR 24-OCT-1997; 97US-0063045.  
XX PR 24-OCT-1997; 97US-0063120.  
XX PR 24-OCT-1997; 97US-0063121.  
XX PR 24-OCT-1997; 97US-0063127.  
XX PR 24-OCT-1997; 97US-0063128.  
XX PR 27-OCT-1997; 97US-0063329.  
XX PR 27-OCT-1997; 97US-0063327.  
XX PR 28-OCT-1997; 97US-0063541.  
XX PR 28-OCT-1997; 97US-0063542.  
XX PR 28-OCT-1997; 97US-0063544.  
XX PR 28-OCT-1997; 97US-0063549.  
XX PR 28-OCT-1997; 97US-0063550.  
XX PR 28-OCT-1997; 97US-0063564.  
XX PR 29-OCT-1997; 97US-0063435.  
XX PR 29-OCT-1997; 97US-0063704.  
XX PR 29-OCT-1997; 97US-0063732.  
XX PR 29-OCT-1997; 97US-0063738.  
XX PR 29-OCT-1997; 97US-0063734.  
XX PR 29-OCT-1997; 97US-0064215.  
XX PR 29-OCT-1997; 97US-0063735.  
XX PR 31-OCT-1997; 97US-0063870.  
XX PR 31-OCT-1997; 97US-0064103.  
XX PR 03-NOV-1997; 97US-0064248.  
XX PR 07-NOV-1997; 97US-0064809.  
XX PR 12-NOV-1997; 97US-0065186.  
XX PR 17-NOV-1997; 97US-0065846.  
XX PR 18-NOV-1997; 97US-0065893.  
XX PR 21-NOV-1997; 97US-0066120.  
XX PR 21-NOV-1997; 97US-0066364.  
XX PR 24-NOV-1997; 97US-0066772.  
XX PR 24-NOV-1997; 97US-0066466.  
XX PR 24-NOV-1997; 97US-0066770.  
XX PR 24-NOV-1997; 97US-0066511.  
XX PR 24-NOV-1997; 97US-0066453.

(GETH ) GENENTECH INC.

Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;

WPI; 1999-229533/19.

N-PSDB; AAX52218.

New isolated human genes and polypeptides used in, e.g. treatment of gastrointestinal ulceration

Claim 12; Fig 11; 320pp; English.

XX AAY13344-403 represent secreted and transmembrane human proteins.  
CC The cDNA sequences are obtained from cDNA libraries, prepared from  
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
CC The encoded polypeptides have specific uses based on their homology to  
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders  
CC associated with the preservation and maintenance of gastrointestinal  
CC mucosa and the repair of acute and chronic mucosal lesions

CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
 CC ulceration and congenital microvillus atrophy), skin diseases associated  
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
 CC potent effects on cell growth and development, diseases related to  
 CC growth or survival of nerve cells including Parkinson's disease,  
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as  
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment  
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an  
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have  
 CC therapeutic applications in wound healing and tissue repair; PRO317 can  
 CC be used for treating problems of the kidney, uterus, endometrium, blood  
 CC vessels, or related tissue, e.g. in the heart of genital tract.  
 XX  
 SQ Sequence 205 AA;

Query Match 100.0%; Score 27; DB 20; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FEFVG 5  
 DB 184 fefvg 188  
 |||||

RESULT 10  
 AAY05279  
 ID AAY05279 standard; Protein; 205 AA.  
 XX  
 AC AAY05279;  
 XX  
 DT 22-JUN-1999 (first entry)  
 XX  
 DE FGF-8 homologue PRO187.  
 XX  
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW EBAF-2; inhibitor; tumour growth; cancer; EGF-like homologue;  
 KW FGF-8 homologue.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9914327-A2.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 10-SEP-1998; 98WO-US18824.  
 XX  
 PR 25-NOV-1997; 97US-0065840.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 18-SEP-1997; 97US-0059117.  
 PR 15-OCT-1997; 97US-0052263.  
 PR 17-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.  
 XX  
 PA (GETH ) GENENTECH INC.

PI Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;  
 PI Roy M, Wood WI;  
 XX  
 DR WPI; 1999-229532/19.  
 DR N-PSDB; AAX28429.  
 XX  
 PT Antibodies against specific proteins overexpressed in tumours  
 XX  
 PS Example 1; Fig 2; 130pp; English.  
 XX  
 CC This sequence represents the FGF-8 homologue PRO187.  
 CC The invention relates to antibodies (Ab) that bind to any of the

CC polypeptides (I) designated PRO187; PRO533; PRO214; PRO240; PRO211;  
 CC PRO230; PRO261; PRO246 or EBAF-2. The Ab, or other agents that inhibit  
 CC expression and/or activity of (I) are used: (i) to inhibit growth of  
 CC tumours; and (ii) as diagnostic/prognostic reagents for detection of  
 CC quantification of (I) in cells or tissues, by standard immunoassays, with  
 CC overexpression being indicative of cancer. For therapeutic use, the Ab  
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
 CC Genes expressing (I), many of which are growth factor homologues, are  
 CC overexpressed in some cases of cancer.  
 XX  
 SQ Sequence 205 AA;

Query Match 100.0%; Score 27; DB 20; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FEFVG 5  
 DB 184 fefvg 188  
 |||||

RESULT 11  
 AAB24392  
 ID AAB24392 standard; Protein; 205 AA.  
 XX  
 AC AAB24392;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE Human PRO187 protein sequence SEQ ID NO:21.  
 XX  
 KW Human; PRO; promotion; inhibition; angiogenesis; cardiovascularisation;  
 KW diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;  
 KW angiogenic; proliferative; cardiant; cardiovascular; antiatherosclerotic;  
 KW cytostatic; gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200032221-A2.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 30-NOV-1999; 99WO-US28313.  
 XX  
 PR 01-DEC-1998; 98WO-US25108.  
 PR 16-DEC-1998; 98US-0112850.  
 PR 12-JAN-1999; 99US-0115554.  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 29-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.  
 XX  
 PA (GETH ) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Hillan KJ, Goddard A;  
 XX Godowski PJ, Gurney AL, Klein RD, Kuo SS, Paoni NF, Smith V;  
 XX Watanabe CK, Williams PM, Wood WI;  
 XX WPI; 2000-412154/35.  
 DR N-PSDB; AAA77525.  
 XX

PT Nucleic acids encoding PRO polypeptides useful for preventing,  
 PT diagnosing and treating disorders of cardiovascular, endothelial or  
 - XX angiogenic disorders in mammals -  
 PS Claim 72; Fig 10; 315pp; English.

XX The present invention describes nucleic acids encoding PRO polypeptides  
 CC useful for preventing, diagnosing and treating disorders of  
 CC cardiovascular, endothelial or angiogenic disorder in mammals by  
 CC modulating cell proliferation, angiogenesis and cardiovascularisation,  
 CC and for identifying agonists and antagonists of these processes. The  
 CC nucleic acids and the proteins they encode may be used in the  
 CC prevention, treatment and diagnosis of diseases associated with  
 CC inappropriate PRO expression such as cardiovascular, endothelial or  
 CC angiogenic disorders in mammals (e.g. atherosclerosis, cancers and  
 CC cardiac hypertrophy). For example, the nucleic acids (NAs) and vectors  
 CC containing them and the PRO polypeptide may be used to treat disorders  
 CC associated with decreased PRO expression. AAA77510 to AAA77721 and  
 CC AAB24388 to AAB24435 represent nucleotide and protein sequences used in  
 CC the exemplification of the present invention.

SQ Sequence 205 AA;

Query Match 100.0%; Score 27; DB 21; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5  
 Db 184 fefvg 188

RESULT 12

AA188567  
 ID AAY88567 standard; Protein; 205 AA.  
 XX  
 AC AAY88567;  
 XX  
 DT 09-AUG-2000 (first entry)  
 XX  
 DE Human PRO187 amino acid sequence.  
 XX  
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;  
 KW cell growth proliferation; fibroblast growth factor-8; ADEPT;  
 KW antibody dependent enzyme mediated prodrug therapy; chromosome 8.  
 XX  
 OS Homo sapiens.

XX WO200015666-A2.

XX 23-MAR-2000.

XX 08-SEP-1999; 99WO-US20594.

XX 10-SEP-1998; 98US-0099803.

XX 10-SEP-1998; 98WO-US18824.

XX (GETH) GENENTECH INC.

XX Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;  
 PI WPI; 2000-271386/23.

XX N-PSDB; AAA30024.

XX New isolated antibodies which bind to specific polypeptides used for  
 PT diagnosis and treatment of neoplastic cell growth and proliferation -

XX Example 1; Fig 2; 200pp; English.

XX This sequence represents a human PRO187 amino acid sequence. PRO187  
 CC shares sequence homology with human fibroblast growth factor-8, and the

CC PRO187 gene is located on chromosome 8. The invention relates to isolated  
 CC antibodies which bind to a polypeptide. The "PRO" polypeptides are  
 CC encoded by genes which are over expressed in the genome of tumour cells.  
 CC Vectors and host cells comprising the nucleic acid encoding the  
 CC antibodies are used in the production of the antibodies. The antibodies  
 CC and nucleic acids encoding them are used for diagnosing a tumour in a  
 CC mammal. The antibodies are used for inhibiting the growth of tumour cells  
 CC and identifying compounds that inhibit a biological or immunological  
 CC activity of and/or expression of a PRO187, PRO533, PRO214, PRO240,  
 CC PRO211, PRO230, PRO261, PRO246 or PRO317 polypeptide. The antibody can be  
 CC used in antibody dependent enzyme mediated prodrug therapy (ADEPT) by  
 CC conjugating the antibody to a prodrug-activating enzyme which converts a  
 CC prodrug to an anti-cancer drug. The antibodies can be fluorescently  
 CC labelled and monitored by light microscopy, flow cytometry or fluorimetry  
 CC for diagnosis and prognosis of tumours.

XX Sequence 205 AA;

Query Match 100.0%; Score 27; DB 21; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5  
 Db 184 fefvg 188

RESULT 13

AAU12308  
 ID AAU12308 standard; Protein; 205 AA.

XX AAU12308;

XX 24-OCT-2001 (first entry)

XX Human PRO187 polypeptide sequence.

XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;  
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
 KW adipocyte; A-peptide; factor VIIA; gene therapy.

XX Homo sapiens.

XX WO200140466-A2.

XX 07-JUN-2001.

XX 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

XX 01-DEC-1999; 99WO-US28634.

XX 02-DEC-1999; 99WO-US28551.

XX 02-DEC-1999; 99WO-US28564.

XX 02-DEC-1999; 99WO-US28565.

XX 09-DEC-1999; 99US-0170262.

XX 16-DEC-1999; 99WO-US30095.

XX 20-DEC-1999; 99WO-US30911.

XX 20-DEC-1999; 99WO-US30999.

XX 30-DEC-1999; 99WO-US31243.

XX 06-JAN-2000; 2000WO-US00277.

XX 11-FEB-2000; 2000WO-US00376.

XX 18-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 22-FEB-2000; 2000WO-US04342.

XX 24-FEB-2000; 2000WO-US04914.

XX 24-FEB-2000; 2000WO-US05004.

XX 01-MAR-2000; 2000WO-US05601.

XX 20-MAR-2000; 2000WO-US07377.

XX 21-MAR-2000; 2000WO-US07532.

XX 30-MAR-2000; 2000WO-US08439.



```

PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 10-NOV-2000; 2000WO-US30873.
PA      (GETH ) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI: 2001-408281/43.
DR N-PSDB; AASZ1380.
DR
XX
PT Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical
XX
XX
PS Claim 12; Fig 27A; 813pp; English.
XX
CC AAU12172-AAU12446 represent novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 205 AA;

```

FT Modified-site 58..62 /note= "amidation site"  
 FT Modified-site 60..64 /note= "CAMP- and CGMP-dependent protein kinase phosphorylation site"  
 FT Modified-site 65..69 /note= "casein kinase II phosphorylation site"  
 FT Modified-site 69..75 /note= "N-myristoylation site"  
 FT Modified-site 89..97 /note= "tyrosine kinase phosphorylation site"  
 FT Modified-site 126..130 /note= "N-glycosylation site"  
 FT Modified-site 188..194 /note= "N-myristoylation site"  
 XX WO200077037-A2.  
 XX PD 21-DEC-2000.  
 XX PF 22-MAY-2000; 2000WO-US14042.  
 XX PR 15-JUN-1999; 99US-0139695.  
 PR 20-JUL-1999; 99US-0145070.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 17-AUG-1999; 99US-0149396.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 07-DEC-1999; 98US-0169495.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 20-MAR-2000; 2000WO-US07377.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 15-MAY-2000; 2000WO-US13358.  
 PR 17-MAY-2000; 2000WO-US13705.  
 XX PA (GETH ) GENENTECH INC.  
 XX PI Ashkenazi AJ, Baker KP, Botstein DA, Desnoyers L, Eaton DL;  
 PI Ferrara N, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;  
 PI Godowski PJ, Gurney AL, Kljavin IJ, Mather JP, Napier MA, Pan J;  
 PI Paoni NF, Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM;  
 PI Wood WI, Zhang Z;  
 XX WPI: 2001-050091/06.  
 DR N-PSDB: AAC87033.  
 XX  
 PT Isolated nucleic acid molecule encoding a PRO polypeptide which is a  
 PT transmembrane polypeptide is useful for gene therapy and identification  
 PT of related polypeptides -  
 XX  
 PS Claim 2; Fig 49; 244pp; English.  
 CC  
 CC The present sequence represents a human secreted and transmembrane  
 CC polypeptide. The specification describes human polypeptides, designated  
 CC PRO196, PRO444, PRO183, PRO185, PRO210, PRO215, PRO217, PRO242, PRO288,  
 CC PRO365, PRO1361, PRO1308, PRO1272, PRO1419, PRO499, PRO7170,  
 CC PRO248, PRO353, PRO1318, PRO1600, PRO9940, PRO533, PRO301, PRO187,  
 CC PRO337, PRO1411, PRO4356, PRO245, PRO285, PRO941, PRO10096, PRO6003,  
 CC PRO6004, PRO350, PRO2830 and PRO6309. The biological activity of cells  
 CC can be modulated with agents that bind to these polypeptides, resulting  
 CC in the death of the cells. The polynucleotides encoding these  
 CC polypeptides are useful in the recombinant production of the  
 CC polypeptides, as a hybridisation probe to screen libraries to isolate

CC homologous sequences, or to map the gene. They may also be used for  
 CC analysing genetic disorders, and to produce transgenic animals which are  
 CC useful for the development and screening of therapeutically useful  
 CC reagents. The polynucleotides can also be used in gene therapy e.g. to  
 CC replace a defective gene.

SQ Sequence 205 AA;

Query Match 100.0%; Score 27; DB 22; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEEVG 5  
 Db 184 fefvg 188  
 |||||

Search completed: March 29, 2002, 14:54:56  
 Job time: 100 sec

---

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	27	100.0	156	3	US-08-916-989B-10	Sequence 10, Appl
2	27	100.0	212	1	US-08-463-965A-2	Sequence 2, Appl
3	27	100.0	212	1	US-08-463-169B-21	Sequence 21, Appl
4	27	100.0	212	3	US-09-103-079-21	Sequence 21, Appl
5	27	100.0	216	2	US-08-821-637-3	Sequence 3, Appl
6	27	100.0	221	3	US-09-081-320-5	Sequence 5, Appl
7	27	100.0	221	3	US-09-081-320-16	Sequence 16, Appl
8	27	100.0	597	6	5171850-2	Patent No. 5171850
9	26	96.3	221	3	US-09-081-320-27	Sequence 27, Appl
10	26	96.3	426	2	US-08-416-870C-10	Sequence 10, Appl
11	24	88.9	10	4	US-08-640-737-3	Sequence 3, Appl
12	24	88.9	41	4	US-08-640-737-41	Sequence 41, Appl
13	24	88.9	41	4	US-08-640-737-42	Sequence 42, Appl
14	24	88.9	220	1	US-08-063-552-10	Sequence 10, Appl
15	24	88.9	220	5	PCT-US93-05704-10	Sequence 10, Appl
16	24	88.9	287	4	US-08-640-737-8	Sequence 8, Appl
17	24	88.9	289	2	US-08-580-545B-4	Sequence 4, Appl
18	24	88.9	289	4	US-09-262-653A-4	Sequence 4, Appl
19	24	88.9	289	4	US-08-640-737-2	Sequence 2, Appl
20	24	88.9	289	4	US-08-867-484A-2	Sequence 2, Appl
21	24	88.9	386	4	US-09-045-284A-2	Sequence 2, Appl
22	24	88.9	465	1	US-08-471-496-9	Sequence 9, Appl
23	24	88.9	465	2	US-08-894-840-9	Sequence 9, Appl
24	24	88.9	465	3	US-09-139-675-9	Sequence 9, Appl
25	24	88.9	470	1	US-08-471-496-2	Sequence 2, Appl
26	24	88.9	470	2	US-08-894-840-2	Sequence 2, Appl
27	24	88.9	470	3	US-09-139-675-2	Sequence 2, Appl

```

RESULT 2
US-08-462-965A-2
; Sequence 2, Application US/08462965A
; Patent No. 5728546
; GENERAL INFORMATION:
; APPLICANT: Greene Ph.D., John M.
; APPLICANT: Gruber Ph.D., Joachim R.
; APPLICANT: Rosen, Craig R.
; TITLE OF INVENTION: Fibroblast Growth Factor 13
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,965A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-439 (PF171)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-462-965A-2

```

```

Query Match 100.0%; Score 27; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5
Db 191 FEFVG 195

```

```

RESULT 3
US-08-462-169B-21
; Sequence 21, Application US/08462169B
; Patent No. 5773252
; GENERAL INFORMATION:
; APPLICANT: John Greene and Craig A. Rosen
; TITLE OF INVENTION: Fibroblast Growth Factor-15
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS

```

```

; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,169B
; FILING DATE: 05 JUN 95
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-441 (PF203)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-462-169B-21

```

```

Query Match 100.0%; Score 27; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5
Db 191 FEFVG 195

```

```

RESULT 4
US-09-103-079-21
; Sequence 21, Application US/09103079A
; Patent No. 6013477
; GENERAL INFORMATION:
; APPLICANT: Greene, John M.
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: Fibroblast Growth Factor 15
; FILE REFERENCE: PF203D1
; CURRENT APPLICATION NUMBER: US/09/103,079A
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 08/462,169
; EARLIER FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-103-079-21

```

```

Query Match 100.0%; Score 27; DB 3; Length 212;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5
Db 191 FEFVG 195

```

```

RESULT 5
US-08-821-847-3
; Sequence 3, Application US/08821637
; Patent No. 5912327
; GENERAL INFORMATION:
; APPLICANT: LI, YULING
; APPLICANT: OELKUCH, MARK
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: METHOD OF PURIFYING CHEMOKINES FROM
; INCLUSION BODIES (AS AMENDED)
; NUMBER OF SEQUENCES: 5

```

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
;; STREET: 1100 NEW YORK AVE., SUITE 600  
;; CITY: WASHINGTON  
;; STATE: DC  
;; COUNTRY: USA  
;; ZIP: 20005  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/821.637  
;; FILING DATE: 20-MAR-1997  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: STEFFE, ERIC K. 36,688  
;; REGISTRATION NUMBER:  
;; REFERENCE/DOCKET NUMBER: 1488.0820000/EKS/KMT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 216 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: protein  
;; US-08-821-637-3

Query Match 100.0%; Score 27; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEEVG 5  
|||||

Db 195 PEEVG 199

RESULT 6  
US-09-081-320-5  
; Sequence 5, Application US/09081320  
; Patent No. 6093544  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Meng, Baozhong  
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,320  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,147  
; FILING DATE: 20-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/069,902  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1722  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

;; FILING DATE: 17-DEC-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Goldman, Michael L.  
;; REGISTRATION NUMBER: 30,727  
;; REFERENCE/DOCKET NUMBER: 19603/1722  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (716) 263-1304  
;; TELEFAX: (716) 263-1600  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 221 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-081-320-5

Query Match 100.0%; Score 27; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEEVG 5  
|||||

Db 11 PEEVG 15

RESULT 7  
US-09-081-320-16  
; Sequence 16, Application US/09081320  
; Patent No. 6093544  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Meng, Baozhong  
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,320  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,147  
; FILING DATE: 20-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/069,902  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1722  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-081-320-16

Query Match 100.0%; Score 27; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
| | | | |  
DB 11 FEFVG 15

RESULT 8

US-09-081-320-16  
Patent No. 5171850  
; APPLICANT: FILMUS, JORGE E.; BUICK, RONALD N.  
; TITLE OF INVENTION: INTESTINAL ONCOPETAL GENE  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/239,084  
; FILING DATE: 31-AUG-1988  
; SEQ ID NO:2:  
; LENGTH: 597  
5171850-2

Query Match 100.0%; Score 27; DB 6; Length 597;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
- | | | | |  
DB 141 FEFVG 145

RESULT 9

US-09-081-320-27  
; Sequence 27, Application US/09081320  
; Patent No. 6093544  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; TITLE OF INVENTION: RUPESTRIS STEM FITTING ASSOCIATED VIRUS  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081.320  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,147  
; FILING DATE: 20-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/069,902  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1722  
; TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-081-320-27

Query Match 96.3%; Score 26; DB 3; Length 221;  
Best Local Similarity 80.0%; Pred. No. 58;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
| | | | |  
DB 11 FEFVG 15

RESULT 10

US-08-416-870C-10  
; Sequence 10, Application US/08416870C  
; Patent No. 5824862  
; GENERAL INFORMATION:  
; APPLICANT: HIYOSHI, TORU  
; APPLICANT: MINE, TOSHIKI  
; APPLICANT: KASAKA, KEISUKE  
; APPLICANT: TYSON, ROBERT HUW  
; APPLICANT: PAGE, ANTHONY MILES JOHN  
; TITLE OF INVENTION: DNA ENCODING ATP-DEPENDENT FRUCTOSE 6-PHOSPHATE  
; TITLE OF INVENTION: 1-PHOSPHOTRANSFERASE ORIGINATING FROM PLANT, RECOMBINANT VE  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALL CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/416,870C  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 760-195P(PCT)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 426 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-416-870C-10

Query Match 96.3%; Score 26; DB 2; Length 426;  
Best Local Similarity 80.0%; Pred. No. 11e-02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
| | | | |  
DB 254 FEFVG 258

RESULT 11  
 US-08-640-737-3  
 ; Sequence 3, Application US/08640737  
 ; Patent No. 6215044  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ARROWSMITH, David A.  
 ; APPLICANT: HELLYER, Susan A.  
 ; APPLICANT: DE SILVA, Jacqueline  
 ; APPLICANT: WHITEMAN, Sally A.  
 ; TITLE OF INVENTION: Tomato Xyloglucan Endo-Transglycosylase  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
 ; STREET: 1100 New York Avenue, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: MS Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/640,737  
 ; FILING DATE: 06-MAY-1996  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB94/02467  
 ; FILING DATE: 10-NOV-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9323225.4  
 ; FILING DATE: 10-NOV-1993  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 10 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: YES  
 ; ANTI-SENSE: NO  
 ; FRAGMENT TYPE: Internal  
 US-08-640-737-3

Query Match 88.9%; Score 24; DB 4; Length 10;  
 Best Local Similarity 80.0%; Pred. No. 7.7;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEVVG 5  
 Db 5 FEFLG 9

RESULT 12  
 US-08-640-737-41  
 ; Sequence 41, Application US/08640737  
 ; Patent No. 6215044  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ARROWSMITH, David A.  
 ; APPLICANT: HELLYER, Susan A.  
 ; APPLICANT: DE SILVA, Jacqueline  
 ; APPLICANT: WHITEMAN, Sally A.  
 ; TITLE OF INVENTION: Tomato Xyloglucan Endo-Transglycosylase  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
 ; STREET: 1100 New York Avenue, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: MS Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/640,737  
 ; FILING DATE: 06-MAY-1996  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB94/02467  
 ; FILING DATE: 10-NOV-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9323225.4  
 ; FILING DATE: 10-NOV-1993  
 ; INFORMATION FOR SEQ ID NO: 41:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 41 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 US-08-640-737-41

Query Match 88.9%; Score 24; DB 4; Length 41;  
 Best Local Similarity 80.0%; Pred. No. 31;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEVVG 5  
 Db 5 FEFLG 9

RESULT 13  
 US-08-640-737-42  
 ; Sequence 42, Application US/08640737  
 ; Patent No. 6215044  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ARROWSMITH, David A.  
 ; APPLICANT: HELLYER, Susan A.  
 ; APPLICANT: DE SILVA, Jacqueline  
 ; APPLICANT: WHITEMAN, Sally A.  
 ; TITLE OF INVENTION: Tomato Xyloglucan Endo-Transglycosylase  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
 ; STREET: 1100 New York Avenue, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: MS Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/640,737  
 ; FILING DATE: 06-MAY-1996  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB94/02467  
 ; FILING DATE: 10-NOV-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9323225.4  
 ; FILING DATE: 10-NOV-1993  
 ; INFORMATION FOR SEQ ID NO: 42:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 41 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 US-08-640-737-42

Query Match 88.9%; Score 24; DB 4; Length 41;  
 Best Local Similarity 80.0%; Pred. No. 31;



Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 Db 5 FEFVG 9

RESULT 14  
 US-08-063-552-10  
 ; Sequence 10, Application US/08063552  
 ; Patent No. 5688936  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Robert H  
 ; TITLE OF INVENTION: Vesicle Membrane Transport Proteins  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheldon & Mak  
 ; STREET: 225 South Lake Avenue, Ninth Floor  
 ; CITY: Pasadena  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 91101  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/063,552  
 ; FILING DATE: 19930514  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Farber, Michael B  
 ; REGISTRATION NUMBER: 32,612  
 ; REFERENCE/DOCKET NUMBER: 9067-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (818) 796-4000  
 ; TELEFAX: (818) 795-6321  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 220 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: Internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Rattus rattus  
 ; US-08-063-552-10

Query Match 88.9%; Score 24; DB 1; Length 220;  
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 Db 178 YEFVG 182

RESULT 15  
 PCT-US93-05704-10  
 ; Sequence 10, Application PC/TUS9305704  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Robert H  
 ; TITLE OF INVENTION: Vesicle Membrane Transport Proteins  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheldon & Mak  
 ; STREET: 225 South Lake Avenue, Ninth Floor  
 ; CITY: Pasadena  
 ; STATE: California  
 ; COUNTRY: USA

ZIP: 91001  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/05704  
 ; FILING DATE: 19930611  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Farber, Michael B  
 ; REGISTRATION NUMBER: 32,612  
 ; REFERENCE/DOCKET NUMBER: 9067-1PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (818) 796-4000  
 ; TELEFAX: (818) 795-6321  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 220 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: Internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Rattus rattus  
 ; PCT-US93-05704-10

Query Match 88.9%; Score 24; DB 5; Length 220;  
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 Db 178 YEFVG 182

Search completed: March 29, 2002, 14:53:37  
 Job time: 21 sec



GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 29, 2002, 14:53:16 ; Search time 12.86 Seconds  
(without alignments)  
29.617 Million cell updates/sec

Title: US-09-608-892-16

Perfect score: 27

Sequence: 1 FEFVG 5

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	131	D83915	hypothetical prote
2	27	100.0	156	A48886	8-oxo-7,8-dihydrog
3	27	100.0	156	I49446	8-oxo-dGTPase - mo
4	27	100.0	198	D83993	NAD(P)H oxidoreduc
5	27	100.0	216	JG5972	fibroblast growth
6	27	100.0	239	C81935	probable ribonucle
7	27	100.0	239	E81169	ribonuclease III N
8	27	100.0	287	B82310	sulfate ABC transp
9	27	100.0	366	A84209	hypothetical prote
10	27	100.0	374	T10165	site-specific DNA-
11	27	100.0	421	S26605	myb-related protei
12	27	100.0	439	S56491	hypothetical prote
13	27	100.0	445	F71033	hypothetical 46.0K
14	27	100.0	448	B75180	antibiotic/peptide
15	27	100.0	508	E54089	probable seed stor
16	27	100.0	511	E84685	glucuronosyltransf
17	27	100.0	527	S15089	phenylalanine--trn
18	27	100.0	552	T28752	development-specif
19	27	100.0	597	A30814	kinase-binding pro
20	27	100.0	645	T39614	hypothetical prote
21	27	100.0	678	B84856	hypothetical prote
22	27	100.0	724	B83342	hypothetical prote
23	27	100.0	764	A84671	thyrotropin-releas
24	27	100.0	1025	I59331	DNA-binding protei
25	27	100.0	1128	DNBEM1	hypothetical prote
26	27	100.0	1857	T50513	probable membrane
27	27	100.0	2493	S45734	DN-cadherin - frui
28	27	100.0	3097	T00021	hypothetical prote
29	26	96.3	75	E86891	

30	26	96.3	157	2	T42324	hypothetical prote
31	26	96.3	179	2	A46447	hydrogenase-relate
32	26	96.3	256	2	B81394	hypothetical prote
33	26	96.3	275	2	F69229	hypothetical prote
34	26	96.3	303	2	E86591	CT580 hypothetical
35	26	96.3	303	2	D72031	CT580 hypothetical
36	26	96.3	352	2	T39065	hypothetical prote
37	26	96.3	381	2	D71981	gdp-d-mannose dehy
38	26	96.3	381	2	D64525	gluconate permease
39	26	96.3	446	2	JC4989	low affinity glucos
40	26	96.3	446	2	E86010	biotin carboxylase
41	26	96.3	501	2	D64453	D12 protein - yea
42	26	96.3	525	2	S57691	DNA polymerase III
43	26	96.3	560	2	D70157	probable membrane
44	26	96.3	674	2	S46092	hypothetical prote
45	26	96.3	877	2	C71492	

## ALIGNMENTS

## RESULT 1

D83915

hypothetical protein BH2124 [Imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000

C:Accession: D83915

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331; 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20263314

A:Accession: D83915

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-131 &lt;STO&gt;

A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05843.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2124

Query Match 100.0%; Score 27; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5

Db 69 FEFVG 73

## RESULT 2

A48886

8-oxo-7,8-dihydroguanosine triphosphatase - human

C:Species: Homo sapiens (man)

C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000

C:Accession: A48886

R:Sakumi, K.; Furuichi, M.; Tsuzuki, T.; Kakuma, T.; Kawabata, S.; Maki, H.; Sekiguch

J. Biol. Chem. 268, 23524-23530, 1993

A:Title: Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP

A:Reference number: A48886; MUID:94043152

A:Accession: A48886

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-156 &lt;SAK&gt;

A:Cross-references: GB:D16581; NID:g2077946; PIDN:BA04013.1; PID:g452589

A:Experimental source: Jurkat T-cell lymphoma

A:Note: sequence extracted from NCBI backbone (NCBI:138927, NCBI:138930)

C:Superfamily: unassigned mut domain proteins; mut domain homology

F:32-66/Domain: mut domain homology &lt;MUT&gt;

Query Match 100.0%; Score 27; DB 2; Length 156;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
|||||  
Db 72 FEVVG 76

## RESULT 3

8'-oxo-dGTPase - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 26-May-2000  
C:Accession: I49446  
R:Kakuma, T.; Nishida, J.  
J. Biol. Chem. 270, 25942-25948, 1995  
A:Title: Mouse MTH1 protein with 8'-oxo-7, 8-dihydro-2'-deoxyguanosine 5'-triphosphatase  
A:Reference number: I49446; MUID:96029697  
A:Accession: I49446  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-156 <RES>  
A:Cross-references: GB:D49956; NID:g1065597; PIDN:BA08711.1; PID:g1065598  
C:Superfamily: unassigned mutt domain proteins; mutt domain homology  
F:32-66/Domain: mutt domain homology <MUTT>

Query Match 100.0%; Score 27; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
-|||||  
Db 72 FEVVG 76

## RESULT 4

NAD(P)H oxidoreductase BH2748 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: D83993  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20263314  
A:Accession: D83993  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <STO>  
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06467.1; GSPDB:GNOC  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2748

Query Match 100.0%; Score 27; DB 2; Length 198;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
|||||  
Db 155 FEVVG 159

## RESULT 5

fibroblast growth factor-17 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: JC5972  
R:Hoshikawa, M.; Ohbayashi, N.; Yonamine, A.; Konishi, M.; Ozaki, K.; Fukui, S.; Itoh, N  
Biochem. Biophys. Res. Commun. 244, 187-191, 1998  
A:Title: Structure and expression of a novel fibroblast growth factor, FGF-17, preferential

A:Reference number: JC5972; MUID:98183421

A:Accession: JC5972

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-216 <HOS>

A:Cross-references: GB:AB009250

Query Match 100.0%; Score 27; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
|||||  
Db 195 FEVVG 199

## RESULT 6

Probable ribonuclease III (EC 3.1.26.3) NMA0888 [imported] - Neisseria meningitidis (C81935)  
C:Species: Neisseria meningitidis  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: C81935  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491  
A:Reference number: A81775; MUID:2022556  
A:Accession: C81935  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-239 <PAR>  
A:Cross-references: GB:AL157959; NID:g7379424; PIDN:CAB84168.1; PID:g737  
A:Experimental source: serogroup A, strain 22491  
C:Genetics:  
A:Gene: rnc; NMA0888  
C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology  
C:Keywords: hydrolase

Query Match 100.0%; Score 27; DB 2; Length 239;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
|||||  
Db 45 FEVVG 49

## RESULT 7

Ribonuclease III NMB0686 [imported] - Neisseria meningitidis (strain MC58 serogroup E E81169)  
C:Species: Neisseria meningitidis  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: E81169  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755  
A:Accession: E81169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-239 <TET>  
A:Cross-references: GB:AE002423; GB:AE002098; NID:g7225913; PIDN:AAF41104.1; PID:g722  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0686  
C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology

Query Match 100.0%; Score 27; DB 2; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 Db 45 FEFGV 49

RESULT 8  
 B82310  
 sulfate ABC transporter, permease protein VC0540 [imported] - Vibrio cholerae (strain N1)  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: B82310  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833  
 A:Accession: B82310  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-287 <HEI>  
 A:Cross-references: GB:AE004140; GB:AE003852; NID:g9654965; PIDN:AAF93708.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0540  
 A:Map position: 1  
 C:Superfamily: maltose transport protein malG

Query Match 100.0%; Score 27; DB 2; Length 287;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 Db 91 FEFGV 95

RESULT 9  
 A84209  
 hypothetical protein Vng0505c [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84209  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: A84209  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-366 <STO>  
 A:Cross-references: GB:AE004437; NID:gl0580109; PIDN:AAG19037.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG0505C

Query Match 100.0%; Score 27; DB 2; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 Db 95 FEFGV 99

RESULT 10  
 T10165  
 site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) NgovII - Neisseria  
 C:Species: Neisseria gonorrhoeae  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: T10165  
 R:Stein, D.C.; Gunn, J.S.; Radlinska, M.; Piekarczyk, A.  
 Gene 157, 19-22, 1995  
 A:Title: Restriction and modification systems of Neisseria gonorrhoeae.  
 A:Reference number: Z16969; MUID:95331562  
 A:Accession: T10165  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-374 <STE>  
 A:Cross-references: EMBL:U43736; NID:gl165244; PIDN:AAA86270.1; PID:gl165245  
 A:Experimental source: strain 1291c  
 C:Genetics:  
 A:Gene: dcmg  
 C:Superfamily: DNA methyltransferase (cytosine-specific)  
 C:Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

Query Match 100.0%; Score 27; DB 2; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 Db 308 FEFGV 312

RESULT 11  
 S26605  
 myb-related protein 1 - garden petunia  
 C:Species: Petunia x hybrida (garden petunia)  
 C>Date: 25-Feb-1994 #sequence\_revision 03-Nov-1995 #text\_change 22-Oct-1999  
 C:Accession: S26605  
 R:Avila, J.; Nieto, C.; Canas, L.; Benito, M.; Paz-Ares, J.  
 submitted to the EMBL Data Library, July 1992  
 A:Description: Petunia hybrida genes related to the maize regulatory C1 gene and to a  
 A:Reference number: S26604  
 A:Accession: S26605  
 A:Molecule type: DNA  
 A:Residues: 1-421 <AVI>  
 A:Cross-references: EMBL:Z13996; NID:g20562; PIDN:CAA78386.1; PID:g20563  
 A:Experimental source: strain v26, developing flowers  
 C:Superfamily: petunia myb-related protein 1; myb DNA-binding repeat homology  
 C:Keywords: DNA binding; duplication; nucleus; transcription regulation  
 F:9-61/Domain: myb DNA-binding repeat homology <MYB1>  
 F:62-112/Domain: myb DNA-binding repeat homology <MYB2>

Query Match 100.0%; Score 27; DB 1; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 Db 278 FEFGV 282

RESULT 12  
 S56491  
 hypothetical 46.0K protein (pepa-gntv intergenic region) - Escherichia coli  
 N:Alternate names: hypothetical protein f439  
 C:Species: Escherichia coli  
 C>Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 20-Aug-1999  
 C:Accession: S56491; D65239  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56491

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-439 <BUR>

A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97162.1; PID:g537107  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: D65239

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-439 <BLAT>

A:Cross-references: GB:AE000497; GB:U00096; NID:g1790711; PIDN:AAC77222.1; PID:g1790716;  
 A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yjgt

C:Superfamily: D-serine permease

Query Match 100.0%; Score 27; DB 2; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5

Db 256 FEFVG 260  
 |||||

#### RESULT 13

F71033 -

hypothetical protein PH1558 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C>Date: 29-Jan-1999 #sequence\_revision 29-Jan-1999 #text\_change 21-Jul-2000

C:Accession: F71033

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki  
 M.; Ohtsuka, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137

A:Accession: F71033

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-445 <RAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30670.1; PID:g3257987

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1558

C:Superfamily: conserved hypothetical protein MTH1070

Query Match 100.0%; Score 27; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5

Db 171 FEFVG 175  
 |||||

#### RESULT 14

B75180

antibiotic/peptide maturation related protein PAB1956 - *Pyrococcus abyssi* (strain Orsay)

C:Species: *Pyrococcus abyssi*

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C:Accession: B75180

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: B75180

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-448 <RAW>

A:Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49521.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1956

C:Superfamily: conserved hypothetical protein MTH1070

Query Match 100.0%; Score 27; DB 2; Length 448;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5

Db 174 FEFVG 178  
 |||||

#### RESULT 15

E64089

probable transport protein HI0736 - *Haemophilus influenzae* (strain Rd KW20)

C:Species: *Haemophilus influenzae*

C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 20-Aug-1999

C:Accession: E64089

R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

A:Reference number: A64000; MUID:95350630

A:Accession: E64089

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 <TIGR>

A:Cross-references: GB:U32757; GB:I42023; NID:g1573740; PIDN:AAC22395.1; PID:g1573742

C:Superfamily: gamma-aminobutyric acid transporter

C:Keywords: transmembrane protein

Query Match 100.0%; Score 27; DB 2; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEFVG 5

Db 151 FEFVG 155  
 |||||

Search completed: March 29, 2002, 14:53:57

Job time: 41 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2002, 14:53:16 ; Search time 10.12 Seconds  
(without alignments)  
18.115 Million cell updates/sec

Title: US-09-608-892-16  
Perfect score: 27  
Sequence: 1 FEFVG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	156	1	P36639 homo sapien
2	27	100.0	156	1	P33688 mus musculus
3	27	100.0	156	1	P33699 rattus norv
4	27	100.0	216	1	O60258 homo sapien
5	27	100.0	216	1	O70627 mus musculus
6	27	100.0	374	1	Q59606 neisseria g
7	27	100.0	439	1	P39344 escherichia
8	27	100.0	508	1	P44849 haemophilus
9	27	100.0	515	1	P41990 caenorhabdi
10	27	100.0	527	1	P36510 rattus norv
11	27	100.0	580	1	P31654 homo sapien
12	27	100.0	597	1	P13265 rattus norv
13	27	100.0	645	1	P78963 schizosacch
14	27	100.0	717	1	O64963 prunus aviu
15	27	100.0	1024	1	O8uku6 homo sapien
16	27	100.0	1025	1	O10836 rattus norv
17	27	100.0	1128	1	P24910 herpesvirus
18	27	100.0	2493	1	P35194 saccharomyc
19	27	100.0	3097	1	O15943 drosophila
20	26	96.3	77	1	P43174 ambrosia ps
21	26	96.3	77	1	MP5B_AMPBS
22	26	96.3	179	1	P43175 ambrosia ps
23	26	96.3	352	1	P31913 alcaligenes
24	26	96.3	446	1	Q92348 schizosacch
25	26	96.3	501	1	P46858 escherichia
26	26	96.3	525	1	Q58626 methanococc
27	26	96.3	525	1	P50076 saccharomyc
28	26	96.3	527	1	O9pwf7 rana rugosa
29	24	88.9	674	1	P38315 saccharomyc
30	24	88.9	160	1	Q21828 caenorhabdi
31	24	88.9	174	1	P25338 saccharomyc
32	24	88.9	247	1	P24806 arabidopsis
33	24	88.9	271	1	O9xym0 drosophila
	24	88.9	283	1	P35694 glycine max

34	24	88.9	290	1	MSME_STRMU	Q00750 streptococc
35	24	88.9	295	1	SM4D_CHICK	Q00665 gallus gall
36	24	88.9	340	1	ERNA_ARTS3	P09891 arthrobacte
37	24	88.9	347	1	UTR2_YEAST	P32623 saccharomyc
38	24	88.9	357	1	FULL1_ACHFU	P35905 achatina fu
39	24	88.9	385	1	CNS1_YEAST	P33313 saccharomyc
40	24	88.9	399	1	PLSX_RHOCA	P30789 rhodobacter
41	24	88.9	402	1	YJ73_YEAST	P39542 saccharomyc
42	24	88.9	413	1	ACKA_SYNT3	P73162 synechocyst
43	24	88.9	413	1	YF71_MYCNY	P55802 mycoplasma
44	24	88.9	426	1	Y064_CHLMU	Q9pin5 chlamydia m
45	24	88.9	426	1	Y962_CHLTR	O84698 chlamydia t

## ALIGNMENTS

RESULT	1
80DP_HUMAN	
ID	80DP_HUMAN
AC	P36639; STANDARD; PRT; 156 AA.
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	7,8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (EC 3.1.6.-) (8-OXO-DGTPASE).
GN	NUDT1 OR MTH1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX	MEDLINE=94043152; PubMed=8226881;
RA	Sakumi K., Furuichi M., Tsuzuki T., Kakuma T., Kawabata S.-I.,
RA	Maki H., Sekiguchi M.;
RT	"Cloning and expression of cDNA for a human enzyme that hydrolyzes
RT	8-oxo-dGTP, a mutagenic substrate for DNA synthesis.";
RL	J. Biol. Chem. 268:23524-23530(1993).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=95229148; PubMed=7713500;
RA	Furuichi M., Yoshida M.C., Oda H., Tajiri T., Nakabeppu Y.,
RA	Tsuzuki T., Sekiguchi M.;
RT	"Genomic structure and chromosome location of the human mutr
RT	homologue gene MTH1 encoding 8-oxo-dgtpase for prevention of A:T to
RT	C:G transversion.";
RL	Genomics 24:485-490(1994).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97362283; PubMed=9211940;
RA	Oda H., Nakabeppu Y., Furuichi M., Sekiguchi M.;
RA	"Regulation of expression of the human MTH1 gene encoding
RT	8-oxo-dgTPase. Alternative splicing of transcription products.";
RL	J. Biol. Chem. 272:17843-17850(1997).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20007875; PubMed=10536140;
RA	Oda H., Taketomi A., Maruyama R., Itoh R., Nishio K., Yakushiji H.,
RA	Suzuki T., Sekiguchi M., Nakabeppu Y.;
RT	"Multi-forms of human MTH1 polypeptides produced by alternative
RT	translation initiation and single nucleotide polymorphism.";
RL	Nucleic Acids Res. 27:4335-4343(1999).
CC	!- FUNCTION: ANTIMUTAGENIC. RESPONSIBLE FOR PREVENTING
CC	MISINCORPORATION OF 8-OXO-DGTP INTO DNA THUS PREVENTING
CC	TRANSVERSIONS.
CC	!- CATALYTIC ACTIVITY: 8-OXO-DGTP + H(2)O -> 8-OXO-DGMP +
CC	PYROPHOSPHATE.
CC	!- PTM: THE N-TERMINUS IS BLOCKED.
CC	!- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its



CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D16581; BAA04013.1; -  
 DR EMBL; D38594; BAA07601.1; -  
 DR EMBL; D38592; BAA07601.1; JOINED.  
 DR EMBL; D38593; BAA07601.1; JOINED.  
 DR EMBL; AB025233; BAA83791.1; -  
 DR EMBL; AB025234; BAA83792.1; -  
 DR EMBL; AB025236; BAA83794.1; -  
 DR EMBL; AB025238; BAA83796.1; -  
 DR PIR; A48886; A48886.  
 DR MIM; 600312; -  
 DR InterPro; IPR000086; NUDIX\_hydrolase.  
 DR Pfam; PF00293; mutt; 1.  
 DR PRINTS; PR00502; MUTTDOMAIN.  
 DR PROSITE; PS00893; NUDIX; 1.  
 KW Hydrolase.  
 FT DOMAIN 37 58 NUDIX BOX.  
 SQ SEQUENCE 156 AA; 17951 MW; B9FB669FF0ACFF5F CRC64;

Query Match 100.0%; Score 27; DB 1; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
 Db 72 FEVVG 76

RESULT 2  
 BODP\_MOUSE STANDARD; PRT; 156 AA.  
 AC P53368; P97795;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE 7,8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (EC 3.1.6.-) (8-OXO-DGTPASE).  
 GN NUDT1 OR MTH1.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=96029697; PubMed=7592783;  
 RA Kakuma T., Nishida J., Tsuzuki T., Sekiguchi M.;  
 RT "Mouse MTH1 protein with 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-  
 RT triphosphatase activity that prevents transversion mutation. cDNA  
 RT cloning and tissue distribution.";  
 RL J. Biol. Chem. 270:25942-25948(1995).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=97166235; PubMed=9013634;  
 RA Igarashi H., Tsuzuki T., Kakuma T., Tomimaga Y., Sekiguchi M.;  
 RT "Organization and expression of the mouse MTH1 gene for preventing  
 RT transversion mutation.";  
 RL J. Biol. Chem. 272:3766-3772(1997).

CC -!- FUNCTION: ANTIMUTAGENIC. RESPONSIBLE FOR PREVENTING  
 CC MISINCORPORATION OF 8-OXO-DGTP INTO DNA THUS PREVENTING A:T TO C:G  
 CC TRANSVERSIONS.  
 CC -!- CATALYTIC ACTIVITY: 8-OXO-DGTP + H(2)O -> 8-OXO-DGMP +  
 CC PYROPHOSPHATE.

CC -!- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D49956; BAA08711.1; -  
 DR EMBL; D88356; BAA19866.1; -  
 DR MGD; MGI:109280; Mth1.  
 DR InterPro; IPR000086; NUDIX\_hydrolase.  
 DR Pfam; PF00293; mutt; 1.  
 DR PRINTS; PR00502; MUTTDOMAIN.  
 DR PROSITE; PS00893; NUDIX; 1.  
 KW Hydrolase.  
 FT DOMAIN 37 58 NUDIX BOX.  
 SQ SEQUENCE 156 AA; 17908 MW; 9E6C12EC2A6DEAB7 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5  
 Db 72 FEVVG 76

RESULT 3  
 BODP\_RAT STANDARD; PRT; 156 AA.  
 AC P53369;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE 7,8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (EC 3.1.6.-) (8-OXO-DGTPASE).  
 GN NUDT1 OR MTH1.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=DONRYU; TISSUE=Spleen;  
 RX MEDLINE=96033912; PubMed=7586133;  
 RA Cai J.P., Kakuma T., Tsuzuki T., Sekiguchi M.;  
 RT "cDNA and genomic sequences for rat 8-oxo-dGTPase that prevents  
 RT occurrence of spontaneous mutations due to oxidation of guanine  
 RT nucleotides.";  
 RL Carcinogenesis 16:2343-2350(1995).

CC -!- FUNCTION: ANTIMUTAGENIC. RESPONSIBLE FOR PREVENTING  
 CC MISINCORPORATION OF 8-OXO-DGTP INTO DNA THUS PREVENTING A:T TO C:G  
 CC TRANSVERSIONS.  
 CC -!- CATALYTIC ACTIVITY: 8-OXO-DGTP + H(2)O -> 8-OXO-DGMP +  
 CC PYROPHOSPHATE.

CC -!- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D49977; BAA08726.1; -  
 DR EMBL; D49980; BAA08727.1; -  
 DR InterPro; IPR000086; NUDIX\_hydrolase.  
 DR Pfam; PF00293; mutt; 1.  
 DR PRINTS; PR00502; MUTTDOMAIN.  
 DR PROSITE; PS00893; NUDIX; 1.

KW Hydrolase.  
 FT DOMAIN 37 58 NUDIX BOX.  
 SQ SEQUENCE 156 AA; 18018 MW; 60AFB6522CB03E18 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 |||||  
 Db 72 FEFVG 76

RESULT 4  
 FGFH\_HUMAN  
 ID FGFH\_HUMAN STANDARD; PRT; 216 AA.  
 AC O60258;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE FIBROBLAST GROWTH FACTOR-17 PRECURSOR (FGF-17).  
 GN FGF17.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=98183421; PubMed=9514906;  
 RA Hoshikawa M., Ohbayashi N., Yonamine A., Konishi M., Ozaki K.,  
 RA Fukui S., Itoh N.;  
 RT "Structure and expression of a novel fibroblast growth factor, FGF-17,  
 RT preferentially expressed in the embryonic brain.";  
 RL Biochem. Biophys. Res. Commun. 244:187-191(1998).  
 CC -1- FUNCTION: MAY BE A SIGNALING MOLECULE IN THE INDUCTION AND  
 CC PATTERNING OF THE EMBRYONIC BRAIN.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE EMBRYONIC  
 CC BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: DETECTED IN EMBRYOS AT E14.5, BUT NOT AT  
 CC E10.5 AND E19.5. PREFERENTIALLY EXPRESSED IN THE NEUROEPITHELIA OF  
 CC THE ISTHUS AND SEPTUM OF THE EMBRYONIC BRAIN AT E14.5.  
 CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB009249; BAA25429.1; -  
 CC MIM; 603725; -  
 CC InterPro; IPR002209; HBGFG\_FGF.  
 CC InterPro; IPR002348; ILL\_HBGF.  
 CC Pfam; PF00167; FGF; 1.  
 CC PRINTS; PR00262; ILLHBGF.  
 CC ProDom; PD000831; HBGFG\_FGF; 1.  
 CC SMART; SM00442; FGF; 1.  
 CC PROSITE; PS00247; HBGFG\_FGF; 1.  
 CC Growth factor; Signal.  
 CC SIGNAL 1 22 POTENTIAL.  
 CC CHAIN 23 216 FIBROBLAST GROWTH FACTOR-17.  
 CC CARBOHYD 137 137 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SEQUENCE 216 AA; 24891 MW; 2EE0288675220F4C CRC64;

Query Match 100.0%; Score 27; DB 1; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 |||||  
 Db 195 FEFVG 199

RESULT 5  
 FGFH\_MOUSE  
 ID FGFH\_MOUSE STANDARD; PRT; 216 AA.  
 AC 070627;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE FIBROBLAST GROWTH FACTOR-17 PRECURSOR (FGF-17).  
 GN FGF17.  
 OS Mus musculus (Mouse), and  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090, 10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Embryo;  
 RX MEDLINE=98183421; PubMed=9514906;  
 RA Hoshikawa M., Ohbayashi N., Yonamine A., Konishi M., Ozaki K.,  
 RA Fukui S., Itoh N.;  
 RT "Structure and expression of a novel fibroblast growth factor, FGF-17,  
 RT preferentially expressed in the embryonic brain.";  
 RL Biochem. Biophys. Res. Commun. 244:187-191(1998).  
 CC -1- FUNCTION: MAY BE A SIGNALING MOLECULE IN THE INDUCTION AND  
 CC PATTERNING OF THE EMBRYONIC BRAIN.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN EMBRYONIC BRAIN, MOSTLY IN THE  
 CC ISTHUS CEREBELLAR NEUROEPITHELIUM AND SEPTUM NEUROEPITHELIUM, AND  
 CC IN ALL ADULT TISSUES.  
 CC -1- DEVELOPMENTAL STAGE: IN RAT, EXPRESSED AT HIGH LEVEL IN THE BRAIN  
 CC EMBRYO AT E14.5. EXPRESSED AT LOWER LEVEL IN THE BRAIN EMBRYO AT  
 CC E10.5 AND E19.5.  
 CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB009250; BAA25430.1; -  
 CC EMBL; AB008682; BAA25426.1; -  
 CC MGI; MGI:1202401; Fgf17.  
 CC InterPro; IPR002209; HBGFG\_FGF.  
 CC InterPro; IPR002348; ILL\_HBGF.  
 CC Pfam; PF00167; FGF; 1.  
 CC PRINTS; PR00262; ILLHBGF.  
 CC ProDom; PD000831; HBGFG\_FGF; 1.  
 CC SMART; SM00442; FGF; 1.  
 CC PROSITE; PS00247; HBGFG\_FGF; 1.  
 CC Growth factor; Signal.  
 CC SIGNAL 1 22 POTENTIAL.  
 CC CHAIN 23 216 FIBROBLAST GROWTH FACTOR-17.  
 CC CARBOHYD 137 137 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SEQUENCE 216 AA; 24924 MW; 2EE94BDF75220F4C CRC64;

Query Match 100.0%; Score 27; DB 1; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
 |||||  
 Db 195 FEFVG 199

RESULT 6  
 MTF7\_NEIGO  
 ID MTF7\_NEIGO STANDARD; PRT; 374 AA.



```

RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: PUTATIVE SODIUM-DEPENDENT TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32757; AAC22395.1;
DR TIGR; HI0736;
DR InterPro; IPR000175; Na_neurotran_symport.
DR Pfam; PF0209; SNF; 1.
DR PROSITE; PS00610; NA_NEUROTRAN_SYM_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYM_2; FALSE_NEG.
DR PROSITE; PS0267; NA_NEUROTRAN_SYM_3; 1.
KW Hypothetical protein; Transport; Transmembrane; Symport;
KW Complete proteome.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 469 489 POTENTIAL.
SQ SEQUENCE 508 AA; 55177 MW; 5CC50526DE854BB9 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEVVG 5
Db 151 FEVVG 155
|||||

RESULT 9
APX1-CAEEL STANDARD; PRT; 515 AA.
AC P41990;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE APX-1 PROTEIN PRECURSOR.
GN APX-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```

OX NCBI_TaxID-6239;
RP [1]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-94208066; PubMed-8156602;
RA Mello C.C., Draper B.W., Priess J.R.;
RT "The maternal genes apx-1 and glp-1 and establishment of
RT dorsal-ventral polarity in the early C. elegans embryo.";
RL Cell 77:95-106(1994).
CC -1- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE BLASTOMER CELL.
CC ABP FATE. CONTRIBUTES TO THE ESTABLISHMENT OF THE DORSAL-VENTRAL
CC AXIS IN THE EARLY C.ELEGANS EMBRYO.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 4 COMPLETE AND 1 INCOMPLETE EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07628; AAA17738.1;
DR HSSP; P00740; LIYA.
DR InterPro; IPR001774; DSL.
DR Pfam; PF01414; DSL; 1.
DR Pfam; PF00008; EGF; 4.
DR SMART; SM00051; DSL; 1.
DR SMART; SM00181; EGF; 4.
DR PROSITE; PS00022; EGF_1; 4.
DR PROSITE; PS01186; EGF_2; 3.
KW Differentiation; Repeat; Transmembrane; EGF-like domain;
KW Glycoprotein; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 515 APX-1 PROTEIN.
FT DOMAIN 27 392 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 393 413 POTENTIAL.
FT DOMAIN 414 515 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 173 205 EGF-LIKE 1.
FT DOMAIN 203 238 EGF-LIKE 2.
FT DOMAIN 240 280 EGF-LIKE 3.
FT DOMAIN 284 322 EGF-LIKE 4.
FT DOMAIN 325 349 EGF-LIKE 5 (INCOMPLETE).
FT DISULFID 177 187 BY SIMILARITY.
FT DISULFID 181 193 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 207 218 BY SIMILARITY.
FT DISULFID 213 226 BY SIMILARITY.
FT DISULFID 228 237 BY SIMILARITY.
FT DISULFID 244 256 BY SIMILARITY.
FT DISULFID 250 268 BY SIMILARITY.
FT DISULFID 270 279 BY SIMILARITY.
FT DISULFID 288 300 BY SIMILARITY.
FT DISULFID 294 310 BY SIMILARITY.
FT DISULFID 312 321 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 515 AA; 55818 MW; F41192A5268C24D9 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 515;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FEVVG 5
Db 272 FEVVG 276
|||||

RESULT 10
UDAL-RAT

```

```

ID UDAL_RAT STANDARD; PRT; 527 AA.
AC P36510;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UDP-GLUCURONOSYLTRANSFERASE 2A1 PRECURSOR, MICROSOMAL (EC 2.4.1.17)
DE (UDPGT) (UGT-OLF).
GN UGT2A1 OR UGT2A-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91156050; PubMed=1900353;
RA Lazar D., Zupko K., Poria Y., Nef P., Lazarovits J., Horn S.,
RA Khem M., Lancel D.;
RT "Odorant signal termination by olfactory UDP glucuronosyl
RT transferase";
RL Nature 349:790-793(1991).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDOGENOUS COMPOUNDS. THIS ISOFORM IS ACTIVE ON ODORANTS AND
CC SEEMS TO BE INVOLVED IN OLFACTION; IT COULD HELP CLEAR LIPOPHILIC
CC ODORANT MOLECULES FROM THE SENSORY EPITHELIUM.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR -> UDP + ACCEPTOR
CC BETA-D-GLUCURONOSIDE.
CC -1- TISSUE SPECIFICITY: OLFACTORY EPITHELIUM.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57565; CAA040797.1; -
DR InterPro: IPR002213; UDPGT.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome; Olfaction.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 527 UDP-GLUCURONOSYLTRANSFERASE 2A1.
FT TRANSMEM 491 507 POTENTIAL.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 527 AA; 59915 MW; 6A32A9B56E8E8DE CRC64;

Query Match 100.0%; Score 27; DB 1; Length 527;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEVVG 5
Db 272 FEVVG 276

RESULT 11
GPC3_HUMAN
ID GPC3_HUMAN STANDARD; PRT; 580 AA.
AC P51654;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE GLYCOPICAN-3 PRECURSOR (Last annotation update)
DE GLYCOPICAN-3 PRECURSOR (INTESTINAL PROTEIN OCI-5) (GPR2-2) (MXR7).
GN GPC3 OR OCI5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97179481; PubMed=9021160;
RA Shen T., Sonoda G., Hamid J., Li M., Filmus J., Buick R.N.,
RA Testa J.R.;
RT "Mapping of the Simpson-Golabi-Behmel overgrowth syndrome gene (GPC3)
RT to chromosome X in human and rat by fluorescence in situ
RT hybridization.";
RL Mamm. Genome 8:72-72(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96172821; PubMed=8589713;
RA Pilla G., Hughes-Benzle R.M., Mackenzie A., Baybayan P., Chen E.Y.,
RA Huber R., Nerl G., Cao A., Forabosco A., Schlessinger D.;
RT "Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel
RT overgrowth syndrome.";
RL Nat. Genet. 12:241-247(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480715; PubMed=9339360;
RA Chen E.Y., Crispont L., Mazzarella R., Chen C.N., Su Y., Shizuya H.,
RA Chen E.Y., Cao A., Pilla G.;
RT "Analysis of exon/intron structure and 400 kb of genomic sequence
RT surrounding the 5'-promoter and 3'-terminal ends of the human glypican
RT 3 (GPC3) gene.";
RL Genomics 45:48-58(1997).
RN [5]
RP SEQUENCE OF 1-112 FROM N.A.
RA Isherwood J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 114-430 FROM N.A.
RA Grahnam D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 431-524 FROM N.A.
RA McDougall R.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. MAY BE
CC INVOLVED IN THE SUPPRESSION/MODULATION OF GROWTH IN THE
CC PREDOMINANTLY MESODERMAL TISSUES AND ORGANS. MAY PLAY A ROLE IN
CC THE MODULATION OF IGF2 INTERACTIONS WITH ITS RECEPTOR AND THEREBY
CC MODULATE ITS FUNCTION. CAN HAVE A POTENTIAL ROLE AS A REGULATOR OF
CC GROWTH AND TUMOR PREDISPOSITION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE LUNG, LIVER AND
CC KIDNEY.
CC -1- DISEASE: DEFECTS IN GPC3 ARE THE CAUSE OF SIMPSON-GOLABI-BEHMEL
CC SYNDROME (SGPS), OR SIMPSON DYSMORPHIA SYNDROME (SDYS) A CONDITION
CC CHARACTERIZED BY PRE- AND POSTNATAL OVERGROWTH WITH VISCERAL AND
CC SKELETAL ANOMALIES.
CC -1- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50410; AAA93471.1; -
DR EMBL; L47125; AAA98132.1; -
DR EMBL; L47176; AAB58806.1; -
DR EMBL; L47124; AAA98131.1; -

```

DR EMBL; 237987; CAA86069.1; -  
 DR EMBL; AF003529; AAB87062.1; -  
 DR EMBL; AL008712; CAA15490.1; -  
 DR EMBL; 295570; CAB16747.1; -  
 DR EMBL; AL009174; CAA15663.1; -  
 DR MIM; 300037; -  
 DR MIM; 312870; -  
 DR InterPro: IPR001863; Glypican.  
 DR Pfam: PF01153; Glypican; 1.  
 DR PROSITE; PS01207; GLYPICAN; 1.  
 KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 ?  
 FT PROPEP ? 580 REMOVED IN MATURE FORM (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 FT CARBOHYD 509 509 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 SQ SEQUENCE 580 AA; 65562 MW; 19485B76D3CE15FC CRC64;

Query Match 100.0%; Score 27; DB 1; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 DB 142 FEFGV 146

RESULT 12  
 GPC3\_RAT  
 ID GPC3\_RAT STANDARD; PRT; 597 AA.  
 AC P13265;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE GLYPICAN-3 PRECURSOR (INTESTINAL PROTEIN OCI-5).  
 GN GPC3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89039851; PubMed=3185547;  
 RA Filmus J., Church J.G., Buick R.N.;  
 RT "Isolation of a cDNA corresponding to a developmentally regulated  
 transcript in rat intestine."  
 RL Mol. Cell. Biol. 8:4243-4249(1988).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96033054; PubMed=7487896;  
 RA Filmus J., Shi W., Wong Z.M., Wong M.J.;  
 RT "Identification of a new membrane-bound heparan sulphate  
 proteoglycan."  
 RL Biochem. J. 311:561-565(1995).  
 CC -1- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. MAY BE  
 INVOLVED IN THE SUPPRESSION/MODULATION OF GROWTH IN THE  
 PREDOMINANTLY MESODERMAL TISSUES AND ORGANS. MAY PLAY A ROLE IN  
 THE MODULATION OF IGF2 INTERACTIONS WITH ITS RECEPTOR AND THEREBY  
 MODULATE ITS FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: FETAL TISSUES AND ADULT LUNG.  
 CC -1- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>;  
 CRC64;

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M22400; AAA41735.1; -  
 CC PIR; A30814; A30814.  
 DR InterPro: IPR001863; Glypican.  
 DR Pfam: PF01153; Glypican; 1.  
 DR PROSITE; PS01207; GLYPICAN; 1.  
 KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 ?  
 FT PROPEP ? 597 REMOVED IN MATURE FORM (POTENTIAL).  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 494 494 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 FT CARBOHYD 508 508 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 SQ SEQUENCE 597 AA; 67048 MW; C556AD4127685687 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 597;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFGV 5  
 |||||  
 DB 141 FEFGV 145

RESULT 13  
 SKB1\_SCHPO  
 ID SKB1\_SCHPO STANDARD; PRT; 645 AA.  
 AC P78963; O42946;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SKB1 KINASE-BINDING PROTEIN 1.  
 GN SKB1 OR SPBC16H5.11C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97098476; PubMed=8943016;  
 RA Gilbrech M., Yang P., Wang D., Frost J., Polverino A., Cobb M.H.;  
 RT "The highly conserved skb1 gene encodes a protein that interacts with  
 Skb1, a fission yeast Ste20/PAK homolog."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13802-13807(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;  
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INTERACTS WITH PAK1/SBK1  
 CC -1- SIMILARITY: TO YEAST YBR133C AND C.ELEGANS C34E10.5.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>;  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U59684; AAC49571.1; -  
 CC EMBL; AL022104; CAA17909.1; -  
 FT CONFLICT 45 45 F -> L (IN REF. 1).  
 FT CONFLICT 134 138 PTSPM -> SNFTQC (IN REF. 1).  
 FT CONFLICT 195 195 A -> G (IN REF. 1).  
 FT CONFLICT 398 398 G -> A (IN REF. 1).  
 SQ SEQUENCE 645 AA; 73166 MW; 288FEE7BDF61558D CRC64;

Query Match 100.0%; Score 27; DB 1; Length 645;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FEFVG 5  
 DB 45 FEFVG 49

## RESULT 14

ID PALL\_PRUV STANDARD; PRT; 717 AA.  
 AC 06493;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PHENYLALANINE AMMONIA-LYASE 1 (EC 4.3.1.5).  
 GN PALL.  
 OS Prunus avium (Cherry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Rosales; Rosaceae; Prunus.  
 OX NCBI\_TaxID=42229;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. SUMMIT;  
 RA Wiersma P.A., Wu Z.;  
 RT "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe  
 (in) 'Plant Gene Register PG98-184'.  
 RL  
 CC -1- FUNCTION: THIS IS A KEY ENZYME OF PLANT METABOLISM CATALYZING THE  
 CC FIRST REACTION IN THE BIOSYNTHESIS FROM L-PHENYLALANINE OF A WIDE  
 CC VARIETY OF NATURAL PRODUCTS BASED ON THE PHENYLPROPANE SKELETON.  
 CC -1- CATALYTIC ACTIVITY: L-PHENYLALANINE - TRANS-CINNAMATE + NH(3).  
 CC -1- PATHWAY: KEY ENZYME OF PHENYLPROPANOID METABOLISM.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- PTM: CONTAINS AN ACTIVE SITE 4-METHYLIDENE-IMIDAZOLE-5-ONE (MIO),  
 CC WHICH IS FORMED AUTOCATALYTICALLY BY CYCLIZATION AND DEHYDRATION  
 CC OF RESIDUES ALA-SER-GLY (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PAL / HISTIDASE FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF036948; AAC78457.1; -  
 CC InterPro; IPR001106; PAL.  
 CC Pfam; PF00221; PAL; 1.  
 CC PROSITE; PS00488; PALL\_HISTIDASE; 1.  
 CC Lyase; Phenylpropanoid metabolism; Multigene family.  
 FT SITE 203 205 MODIFIED TO FORM 4-METHYLIDENE-IMIDAZOLE-5-ONE (BY SIMILARITY).  
 FT SEQUENCE 717 AA; 77999 MW; B84DF90A0BF60B3 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 717;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FEFVG 5  
 DB 239 FEFVG 243

## RESULT 15

THDE\_HUMAN STANDARD; PRT; 1024 AA.

ID THDE\_HUMAN STANDARD; PRT; 1024 AA.  
 AC 09UKU6;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE THYROTROPIN-RELEASING HORMONE DEGRADING ECOTOENZYME (EC 3.4.19.6) (TRH-  
 DE DEGRADING ECOTOENZYME) (TRH-DE) (TRH-SPECIFIC AMINOPEPTIDASE)  
 DE (THYROLIBERINASE) (PYROGLUTAMYL-PEPTIDASE II) (PAP-II).  
 GN TRHDE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=99421758; PubMed=10491199;  
 RA Schomburg L., Turwitt S., Frescher G., Lohmann D., Horsthemke B.,  
 BAUER K.;  
 RT "Human TRH-degrading ectoenzyme cDNA cloning, functional expression,  
 RT genomic structure and chromosomal assignment.";  
 RL Eur. J. Biochem. 265:415-422(1999).  
 CC -1- FUNCTION: SPECIFIC INACTIVATION OF TRH AFTER ITS RELEASE.  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF THE N-TERMINAL PYROGLUTAMYL GROUP  
 CC FROM PGLU-[1-HIS-XAA TRIPEPTIDES AND PGLU-[1-HIS-XAA-GLY  
 CC TETRAPEPTIDES.  
 CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);  
 CC ALSO KNOWN AS THE PEPN SUBFAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF126372; AAF13141.1; -  
 CC InterPro; IPR001930; Aladiptase.  
 CC Pfam; PF01433; Peptidase M1; 1.  
 CC PRINTS; PR00756; ALADIPTASE.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Amino-peptidase; Zinc; Glycoprotein;  
 CC Transmembrane; Signal-anchor.  
 FT DOMAIN 1 40 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 41 61 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 62 1024 EXTRACELLULAR (POTENTIAL).  
 FT METAL 440 440 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 441 441 BY SIMILARITY.  
 FT METAL 444 444 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT 463 463 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 527 527 PROTON DONOR (POTENTIAL).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 605 605 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 906 906 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 1024 AA; 116999 MW; B82FBB9C93927F6DC CRC64;

Query Match

100.0%; Score 27; DB 1; Length 1024;

Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PEFVG 5  
Db 469 PEFVG 473

Search completed: March 29, 2002, 14:55:13  
Job time: 117 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2002, 14:53:16 ; Search time 22.86 Seconds  
(without alignments)  
31.993 Million cell updates/sec

Title: US-09-608-892-16  
Perfect score: 27  
Sequence: 1 FEFVG 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL17:  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_prodent.\*  
12: sp\_virus.\*  
13: sp\_invertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	27	100.0	131	2 Q9KB12	Q9KB12 bacillus ha
2	27	100.0	179	4 Q9UBM9	Q9UBM9 homo sapien
3	27	100.0	197	4 Q9UBM0	Q9UBM0 homo sapien
4	27	100.0	198	2 Q9K9A1	Q9K9A1 bacillus ha
5	27	100.0	221	12 Q91725	Q91725 grapevine r
6	27	100.0	221	12 Q91902	Q91902 rupestris r
7	27	100.0	239	2 Q9K0C8	Q9K0C8 neisseria s
8	27	100.0	239	2 Q9JVD3	Q9JVD3 neisseria m
9	27	100.0	258	2 Q9L2B4	Q9L2B4 streptomyce
10	27	100.0	262	2 Q9L104	Q9L104 streptomyce
11	27	100.0	287	2 Q9KUT1	Q9KUT1 vibrio chol
12	27	100.0	314	11 Q9DA78	Q9DA78 mus musculu
13	27	100.0	365	2 Q9A475	Q9A475 caulobacter
14	27	100.0	366	1 Q9HRX1	Q9HRX1 halobacteri
15	27	100.0	402	13 Q98SE8	Q98SE8 gallus gall
16	27	100.0	421	10 Q02992	Q02992 petunia hyb
17	27	100.0	445	1 Q59214	Q59214 pyrococcus
18	27	100.0	448	1 Q9V130	Q9V130 pyrococcus
19	27	100.0	449	4 Q9H6S4	Q9H6S4 homo sapien

20	27	100.0	468	10 Q9LTP6	Q9LTP6 arabidopsis
21	27	100.0	486	1 Q9HKZ1	Q9HKZ1 thermoplasma
22	27	100.0	503	5 Q917K2	Q917K2 drosophila
23	27	100.0	509	2 Q9CKU5	Q9CKU5 pasteurella
24	27	100.0	509	10 Q9FT92	Q9FT92 arabidopsis
25	27	100.0	511	10 Q9SK09	Q9SK09 arabidopsis
26	27	100.0	513	5 Q9TXN4	Q9TXN4 caenorhabdi
27	27	100.0	527	4 Q9V4X1	Q9V4X1 homo sapien
28	27	100.0	527	11 Q9ESE4	Q9ESE4 mus musculu
29	27	100.0	530	11 Q9R110	Q9R110 cavia porce
30	27	100.0	534	11 Q9D811	Q9D811 mus musculu
31	27	100.0	552	5 Q9GYS8	Q9GYS8 caenorhabdi
32	27	100.0	579	11 Q9Q2F1	Q9Q2F1 mus musculu
33	27	100.0	604	5 Q9VNU7	Q9VNU7 drosophila
34	27	100.0	605	2 Q9AGJ6	Q9AGJ6 mycobacteri
35	27	100.0	609	2 Q9G084	Q9G084 mycobacteri
36	27	100.0	678	10 Q9SIM9	Q9SIM9 arabidopsis
37	27	100.0	724	2 Q91151	Q91151 pseudomonas
38	27	100.0	764	10 Q9XIN5	Q9XIN5 arabidopsis
39	27	100.0	1078	5 Q9VSS9	Q9VSS9 drosophila
40	27	100.0	1857	10 Q9LEX9	Q9LEX9 arabidopsis
41	26	96.3	53	2 Q54338	Q54338 staphylococ
42	26	96.3	75	2 Q9CDS8	Q9CDS8 lactococcus
43	26	96.3	127	11 Q9CTN7	Q9CTN7 mus musculu
44	26	96.3	143	2 Q9F8Q2	Q9F8Q2 carboxydoth
45	26	96.3	157	9 Q48484	Q48484 bacterioph

## ALIGNMENTS

RESULT 1  
ID Q9KB12 PRELIMINARY; PRT; 131 AA.  
AC Q9KB12;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE BH2124 PROTEIN.  
GN BH2124.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A. 9153;  
RC STRAIN=C-125 / JCM 9153;  
RA MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AF001514; BAB05843.1; --  
KW Complete proteome.

SQ SEQUENCE 131 AA; 14925 MW; 34D82A531AE355EB CRC64;

Query Match 100.0%; Score 27; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||  
Db 69 FEFVG 73

RESULT 2  
ID Q9UBM9 PRELIMINARY; PRT; 179 AA.  
AC Q9UBM9;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 GN MTH1B (P22), MTH1C (P21), MTH1D (P18).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T CELL LEUKEMIA;  
 RX MEDLINE=20007875; PubMed=10536140;  
 RA Oda H., Taketomi A., Maruyama R., Itoh R., Nishioka K., Yakushiji H.,  
 RA Suzuki T., Sekiguchi M., Nakabeppu Y.;  
 RT "Multi-forms of human MTH1 polypeptides produced by alternative  
 translation initiation and single nucleotide polymorphism.";  
 RL Nucleic Acids Res. 27:4335-4343(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T CELL LEUKEMIA;  
 RX MEDLINE=97362283; PubMed=9211940;  
 RA Oda H., Nakabeppu Y., Furuichi M., Sekiguchi M.;  
 RT "Regulation of Expression of the Human MTH1 Gene Encoding 8-oxo-  
 dGTPase.";  
 RL J. Biol. Chem. 272:17843-17850(1997).  
 DR EMBL: AB025239; BAA83797.1; -;  
 DR EMBL: AB025235; BAA83793.1; -;  
 DR EMBL: AB025237; BAA83795.1; -;  
 DR InterPro: IPR000086; NUDIX\_hydrolase.  
 DR Pfam: PF00293; mutt; 1.  
 DR PRINTS: PR00502; NUDIXFAMILY.  
 DR PROSITE: PS00893; NUDIX. 1.  
 SQ SEQUENCE 179 AA; 20296 MW; 7C9F192384F66C61 CRC64;  
  
 Query Match 100.0%; Score 27; DB 4; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 FEFVG 5  
 Db 95 FEFVG 99  
  
 RESULT 3  
 Q9UBM0  
 ID Q9UBM0 PRELIMINARY; PRT; 197 AA.  
 AC Q9UBM0  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE MTH1A-MET83 (P26), MTH1B-MET83 (P22), MTH1C-MET83 (P21), MTH1D-MET83  
 DE (P18).  
 GN MTH1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T CELL LEUKEMIA;  
 RX MEDLINE=20007875; PubMed=10536140;  
 RA Oda H., Taketomi A., Maruyama R., Itoh R., Nishioka K., Yakushiji H.,  
 RA Suzuki T., Sekiguchi M., Nakabeppu Y.;  
 RT "Multi-forms of human MTH1 polypeptides produced by alternative  
 translation initiation and single nucleotide polymorphism.";  
 RL Nucleic Acids Res. 27:4335-4343(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T CELL LEUKEMIA;  
 RX MEDLINE=97362283; PubMed=9211940;  
 RA Oda H., Nakabeppu Y., Furuichi M., Sekiguchi M.;  
 RT "Regulation of Expression of the Human MTH1 Gene Encoding 8-oxo-

RT dGTPase.";  
 RL J. Biol. Chem. 272:17843-17850(1997).  
 DR EMBL: AB025242; BAA83800.1; -;  
 DR EMBL: AB025240; BAA83798.1; -;  
 DR EMBL: AB025241; BAA83799.1; -;  
 DR InterPro: IPR000086; NUDIX\_hydrolase.  
 DR Pfam: PF00293; mutt; 1.  
 DR PRINTS: PR00502; NUDIXFAMILY.  
 DR PROSITE: PS00893; NUDIX. 1.  
 SQ SEQUENCE 197 AA; 22552 MW; 82B25F56D382CE57 CRC64;  
  
 Query Match 100.0%; Score 27; DB 4; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 FEFVG 5  
 Db 113 FEFVG 117  
  
 RESULT 4  
 Q9K9A1  
 ID Q9K9A1 PRELIMINARY; PRT; 198 AA.  
 AC Q9K9A1  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE NAD(P)H OXIDOREDUCTASE.  
 GN BH2748.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=85665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20513582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001516; BAB06467.1; -;  
 DR InterPro: IPR003680; NADHdh\_2.  
 DR Pfam: PF02525; NADHdh\_2; 1.  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 198 AA; 22252 MW; 1B712749120FA926 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 FEFVG 5  
 Db 155 FEFVG 159  
  
 RESULT 5  
 O91725  
 ID O91725 PRELIMINARY; PRT; 221 AA.  
 AC O91725;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE TRIPLE GENE BLOCK.  
 OS grapevine Rupestris stem pitting associated virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Foveavirus.  
 OX NCBI\_TaxID=82570;  
 RN [1]  
 RP SEQUENCE FROM N.A.

```

RA Zhang Y.-P., Uyemoto J.K., Rowhani A.;
RT "Nucleotide sequence and RT-PCR detection of a virus associated with
RL grapevine Rupestris stem-pitting disease.";
DR EMBL; AF026278; AAC62911.1; -;
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR PROSITE; PS00881; PROTEIN_SPLICING; UNKNOWN.1.
SQ SEQUENCE 221 AA; 24467 MW; BBEE3FE0F8B963A CRC64;

Query Match 100.0%; Score 27; DB 12; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FEFVG 5
DB 11 FEFVG 15

RESULT 6
O91902 SEQUENCE FROM N.A. PRT; 221 AA.
ID O91902;
AC O91902; PRELIMINARY;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 24.4 KDA PROTEIN.
OS Rupestris stem pitting associated virus-1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Foveavirus.
OX NCBI_TaxID=81832;
[1]
RA Meng B., Pang S.-Z., Forsline P.L., McFerson J.R., Gonsalves D.;
RX MEDLINE=98378060; PubMed=9714258;
RA Meng B., Pang S.-Z., Forsline P.L., McFerson J.R., Gonsalves D.;
RT "Nucleotide sequence and genome structure of grapevine rupestris stem
RT pitting associated virus-1 reveal similarities to apple stem pitting
RL virus.";
RL J. Gen. Virol. 79:0-0(0).
[2]
RA SEQUENCE FROM N.A.
RX MEDLINE=98378060; PubMed=9714258;
RA Meng B., Pang S.-Z., Forsline P.L., McFerson J.R., Gonsalves D.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057136; AAC35499.1; -;
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR PROSITE; PS00881; PROTEIN_SPLICING; UNKNOWN.1.
SQ SEQUENCE 221 AA; 24437 MW; B3511EB101706291 CRC64;

Query Match 100.0%; Score 27; DB 12; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FEFVG 5
DB 11 FEFVG 15

RESULT 7
O9K0C8 SEQUENCE FROM N.A. PRT; 239 AA.
ID O9K0C8;
AC O9K0C8; PRELIMINARY;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RIBONUCLEASE III.
GN NMB0686
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;

```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Saunders N.J., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Gili J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AEO02423; AAF41104.1; -;
DR TIGR; NMB0686; -;
DR InterPro; IPR001159; DS_RBD.
DR InterPro; IPR000999; RNase_3.
DR Pfam; PF00035; dsrm; 1.
DR Pfam; PF00636; Ribonuclease_3; 1.
DR SMART; SM00358; DSRM; 1.
DR SMART; SM00535; RIBOC; 1.
DR PROSITE; PS50137; DS_RBD; 1.
DR PROSITE; PS50142; RNase_3_2; 1.
KW Complete proteome.
SQ SEQUENCE 239 AA; 26853 MW; 54D0F0473049607C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FEFVG 5
DB 45 FEFVG 49

RESULT 8
O9JVD3 SEQUENCE FROM N.A. PRT; 239 AA.
ID O9JVD3 PRELIMINARY;
AC O9JVD3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE RIBONUCLEASE III (EC 3.1.26.3).
GN RNC OR NMA0888.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RA MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84168.1; -;
DR InterPro; IPR001159; DS_RBD.
DR InterPro; IPR000999; RNase_3.
DR Pfam; PF00035; dsrm; 1.
DR Pfam; PF00636; Ribonuclease_3; 1.
DR SMART; SM00358; DSRM; 1.
DR SMART; SM00535; RIBOC; 1.
DR PROSITE; PS50137; DS_RBD; 1.
DR PROSITE; PS50142; RNase_3_2; 1.
KW Hydrolase; Complete proteome.

```

SQ SEQUENCE 239 AA; 26841 MW; E8772B870A940CDF CRC64;

Query Match 100.0%; Score 27; DB 2; Length 239;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5

|||||

Db 45 FEFVG 49

RESULT 9

ID Q9L2B4 PRELIMINARY; PRT; 258 AA.

AC Q9L2B4; the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

GN SC8F4.10C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL137242; CAB70636.1; -

KW Hypothetical protein.

SQ SEQUENCE 258 AA; 27101 MW; 896A77EC95DA1938 CRC64;

Query Match

Best Local Similarity 100.0%; Score 27; DB 2; Length 258;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5

|||||

Db 196 FEFVG 200

RESULT 10

Q9L104

ID Q9L104 PRELIMINARY; PRT; 262 AA.

AC Q9L104;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE PUTATIVE HYDROLASE.

GN SCL6.12C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL159139; CAB76877.1; -

DR InterPro; IPR003010; CN\_hydrolase.

DR Pfam; PF00795; CN\_hydrolase; 1.

DR PROSITE; PS01227; UPF0012; 1.

KW Hydrolase.

SQ SEQUENCE 262 AA; 27723 MW; 896D63A846D0ED31 CRC64;

Query Match

Best Local Similarity 100.0%; Score 27; DB 2; Length 262;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5

|||||

Db 205 FEFVG 209

RESULT 11

Q9KUI1

ID Q9KUI1 PRELIMINARY; PRT; 287 AA.

AC Q9KUI1;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE VULPATE ABC TRANSPORTER, PERMEASE PROTEIN.

GN VC0340.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.;

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.;

RA Gilli S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.;

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.;

RA McDonald L., Otterback T., Fleischmann R.D., Nierman W.C., White O.;

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.;

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

cholerae.";

RL Nature 406:477-483(2000).

CC -1- FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT

SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE

SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-

PROTEIN-DEPENDENT TRANSPORT SYSTEMS.

DR EMBL; AF004140; AAF93708.1; -

DR TIGR; VC0540; -

DR InterPro; IPR000515; BPD transp.

DR Pfam; PF00528; BPD transp. 1.

DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER; 1.

DR Complete proteome; Transmembrane; Transport.

KW

SO SEQUENCE 287 AA; 32032 MW; EC94D3E9F951A0C2 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 287;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||

Db 91 FEFVG 95

RESULT 12

Q9DA78 PRELIMINARY; PRT; 314 AA.

ID Q9DA78

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE 1700018F24RIK PROTEIN.

GN 1700018F24RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TESTIS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Mateu Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).

DR EMBL; AK006086; BAB24403.1;

DR MGD; MGI:1916646; 1700018F24RIK.

DR InterPro; IPR000651; RasGEFN.

DR SMART; SM00229; RasGEFN.1.

SO SEQUENCE 314 AA; 35494 MW; 9176D7288CA3524F CRC64;

Query Match 100.0%; Score 27; DB 11; Length 314;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||

Db 60 FEFVG 64

RESULT 13

Q9A475 PRELIMINARY; PRT; 365 AA.

ID Q9A475

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN CC2968.

GN CC2968.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Pótocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus."

DR Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005960; AAK24930.1;

DR TIGR; CC2968;

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 365 AA; 40174 MW; 78A4CD4E77C06DE0 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5  
|||||

Db 294 FEFVG 298

RESULT 14

Q9HRX1 PRELIMINARY; PRT; 366 AA.

ID Q9HRX1

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE VNG0505C.

GN Halobacterium sp. (strain NRC-1).

OS Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;

OC Halobacterium.

OX NCBI\_TaxID=64091;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20504483; PubMed=11016950;

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,

RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,

RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,

RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,

RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,

RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;

RT "Genome sequence of Halobacterium species NRC-1."

RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

DR EMBL; AE005003; AAG19037.1;

DR InterPro; IPR00644; CBS.

DR Pfam; PF00571; CBS; 2.

DR SMART; PF01595; DUF21; 1.

DR SMART; SM00116; CBS; 2.

KW Complete proteome.

SO SEQUENCE 366 AA; 39573 MW; F09F6F97D14EFE49 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 366;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FEFVG 5

DB 95 FEFVG 99  
|||||

RESULT 15  
Q98SE8  
ID Q98SE8 PRELIMINARY; PRT: 402 AA.  
AC Q98SE8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE UBIQUITIN PROTEIN LIGASE (FRAGMENT).  
GN UBE3B.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lemax M.I., Huang L., Cho Y., Gong T.L., Altschuler R.A.;  
RT "Differential display and gene arrays to examine auditory  
RT plasticity";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF245511; AAG53076.1; -  
DR EMBL; AF251045; AAK28418.1; -  
KW Ligase.  
FT NON\_TER  
FT 1  
SO SEQUENCE 402 AA; 45534 MW; 384C9ADB1ED315E7 CRC64;

Query Match 100.0%; Score 27; DB 13; Length 402;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FEFVG 5  
Db 109 FEFVG 113  
|||||

Search completed: March 29, 2002, 14:54:26  
Job time: 70 sec